

PCTORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international

DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : C12Q 1/68, C12N 15/56, 1/21	A1	(11) Numéro de publication internationale: WO 98/01579 (43) Date de publication internationale: 15 janvier 1998 (15.01.98)
(21) Numéro de la demande internationale: PCT/FR97/01133 (22) Date de dépôt international: 25 juin 1997 (25.06.97) (30) Données relatives à la priorité: 96/08337 4 juillet 1996 (04.07.96) FR (71) Déposant (pour tous les Etats désignés sauf US): CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE [FR/FR]; 3, rue Michel-Ange, F-75016 Paris (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): GRANGER, Michèle [FR/FR]; 24, rue du Gloeckelsberg, F-67113 Blasheim (FR). SCHNARR, Manfred [FR/FR]; 24, rue du Gloeckelsberg, F-67113 Blasheim (FR). (74) Mandataires: GROSSET-FOURNIER, Chantal etc.; Grosset-Fournier & Demachy S.à.r.l., 103, rue La Fayette, F-75481 Paris Cedex 10 (FR).		(81) Etats désignés: JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Publiée <i>Avec rapport de recherche internationale.</i>
(54) Title: SYSTEM FOR DETECTING PROTEIN-PROTEIN INTERACTIONS (54) Titre: SYSTEME DE DETECTION DES INTERACTIONS PROTEINE-PROTEINE (57) Abstract <p>The invention concerns the use of a DNA sequence comprising: a nucleotide sequence coding for an indicator protein (indicator gene); a promoter; a mutated operator; the said nucleotide sequence being under the control of the said promoter comprising the said mutated operator containing two sequences and deriving from an operator containing two inverted repeat sequences (palindrome), this mutated operator being such that one of its sequences is mutated (mutation 1), while the other is for instance wild. The transcription of the indicator gene is repressed by controlling proteins in the form of heterodimers via their DNA binding sites, each monomer of these heterodimers being such that it recognises specifically one the sequence comprising mutation 1, the other the wild sequence, in particular for detecting interactions between two proteins, these interactions being revealed when there is repression of the transcription of the said indicator gene, this repression resulting from the formation of heterodimers.</p> (57) Abrégé <p>La présente invention concerne l'utilisation d'une séquence d'ADN comprenant: une séquence nucléotidique codant pour une protéine indicatrice (gène indicateur), un promoteur, un opérateur muté, ladite séquence nucléotidique étant sous le contrôle du susdit promoteur comprenant le susdit opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est par exemple à l'état sauvage, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, notamment pour détecter des interactions entre deux protéines, ces interactions étant révélées lorsqu'il y a répression de la transcription du susdit gène indicateur, cette répression résultant de la formation d'hétérodimères.</p>		

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brazil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

SYSTEME DE DETECTION DES INTERACTIONS PROTEINE-PROTEINE

L'invention a pour objet un système de détection des interactions protéine-protéine, et plus particulièrement des séquences d'ADN pour la mise en oeuvre de ce système, ainsi que des hôtes cellulaires contenant lesdites séquences d'ADN.

Les interactions protéine-protéine jouent un rôle fondamental chez les organismes vivants supérieurs tant au niveau du développement normal de l'organisme qu'au niveau de certaines situations pathologiques. Les méthodes biochimiques et biophysiques existantes exigent d'une part que les différents partenaires aient été préalablement identifiés et, d'autre part, nécessitent de grandes quantités de protéines (donc leur clonage et leur purification). Il est donc indispensable de disposer d'outils permettant de réaliser l'étape initiale, c'est-à-dire permettant d'identifier et de cloner des protéines interagissant entre elles. De tels systèmes ont été récemment développés en utilisant la levure comme hôte et l'activation de la transcription d'un gène spécifique comme "sonde" (voir Fields et Sternglanz, 1994). Cependant, quoique déjà bien utilisé, ce système présente certaines limites.

L'étude de ces interactions protéine-protéine peut se faire en utilisant des techniques biochimiques et biophysiques *in vitro*. Mais celles-ci ne sont concevables que lorsque les protéines sont clonées et peuvent être produites et purifiées en grande quantité. D'autre part, elles ne permettent pas d'identifier de nouveau partenaire d'interaction.

Une alternative attrayante aux méthodes biochimiques consiste en une utilisation de banques d'expression d'ADNc (reflétant l'ensemble des ARN messagers) ou d'ADN génomique. L'avantage de ces stratégies réside dans le fait que les gènes des protéines ainsi identifiées sont immédiatement disponibles (pour une comparaison de ces méthodes voir Phizicky et Fields, 1995). Parmi les méthodes s'appuyant sur l'utilisation de banques d'expression, on peut distinguer celles qui utilisent une sélection *in vitro* (comme la révélation de plages de lyse par une protéine purifiée et marquée radioactivement interagissant avec une protéine inconnue) de celles utilisant une sélection *in vivo*. Dans ce dernier domaine, des progrès sensibles ont pu être réalisés grâce à la technique du "double-hybride" dans la levure (Ma et Ptashne, 1988; Fields et Song, 1989; Chien *et al.*, 1991; Gyuris *et al.*, 1993; Durfee *et al.*, 1993; Vojtek *et al.*, 1993; Le Douarin *et al.*, 1995).

Dans le système développé par Fields et Song (1989), permettant d'identifier *in vivo* chez la levure des protéines interagissant entre elles, une protéine X connue est greffée sur un domaine de fixation à l'ADN (Gal4 ou LexA) et on cherche une deuxième protéine Y (greffée sur un domaine de l'activation ou "AD-activation domain" de la transcription) capable d'interagir avec la protéine X et de provoquer ainsi l'activation de la transcription. Dans une revue, Fields S. et Sternglanz R. ("The two-hybrid system: an assay for protein-protein interaction" (1994) TIG, 10:286-292) discutent les possibilités et les limites de cette méthode. Une des limites est que les deux protéines hybrides doivent parvenir dans le noyau de la cellule. L'autre difficulté majeure consiste à identifier des protéines Y qui interagissent avec une protéine X pourvue d'un domaine d'activation de la transcription, puisque, dans ce cas, la protéine cible (Gal4-X ou LexA-X) donnera un signal d'activation constitutif en absence de toute protéine Y.

S'agissant de Gal4 comme support de fixation à l'ADN, elle a été initialement mise au point en utilisant le domaine de fixation à l'ADN ou DBD ("DNA binding domain") de la protéine Gal4 (Ma et Ptashne, 1988; Fields et Song, 1989). Gal4 est une protéine modulaire de 881 acides aminés possédant au moins deux domaines fonctionnels. Un domaine de fixation à l'ADN comprenant les acides aminés 1-94 et un domaine d'activation de la transcription. La fonction d'activation est en fait répartie sur trois segments différents de la protéine, à savoir 94-106, 148-196 et 768-881 (Marmorstein *et al.*, 1992). La structure de Gal4₁₋₆₅ impliquée dans un complexe avec l'ADN a été déterminée par cristallographie aux rayons X (Marmorstein *et al.*, 1992) tandis que la structure du domaine d'activation reste très mal définie.

Ainsi, Fields et Song (1989) ont greffé, d'une part la protéine SNF1 (une sérine/thréonine kinase) sur Gal4₁₋₁₄₇, et d'autre part la protéine SNF4 sur le domaine d'activation 768-881 de Gal4. Aucune des deux protéines hybrides, Gal4₁₋₁₄₇-SNF1 ou SNF4-Gal4₇₆₈₋₈₈₁ seules n'étaient capable d'activer la transcription du gène reporter *lacZ* placé sous le contrôle du promoteur GAL1. Par contre la coexpression des deux protéines hybrides a abouti à une activation de la transcription notable (180 unités), bien que nettement inférieure à l'activation observée avec la protéine Gal4 entière (4000 unités). Cette différence est imputable, d'une part à l'absence du domaine d'activation comprenant les acides aminés 148-196 et, d'autre part, au fait que l'interaction entre SNF1 et SNF4 ne remplace pas complètement une liaison covalente entre les deux domaines de Gal4. Une fusion directe entre Gal4₁₋₁₄₇ et Gal4₇₆₈₋₈₈₁

provoque en effet seulement une réduction d'un facteur 2 par rapport à Gal4 entier (Ma et Ptashne, 1987a).

Depuis, des dizaines de couples de protéines X,Y ont été testées. De nombreuses combinaisons sont capables de restituer un activateur fonctionnel de la transcription (pour une revue, voir Bartel *et al.*, 1993).

Outre l'analyse de l'interaction d'un couple de protéines connues, la technique du double-hybride permet également l'identification d'un partenaire inconnu du moment où on est en possession du gène de la protéine cible X. L'équipe de Stanley Fields a été la première à exploiter cette possibilité (Chien *et al.*, 1991) en créant des fusions entre le domaine d'activation 768-881 de Gal4 et des séquences protéiques codées par des fragments d'ADN d'une banque d'ADN génomique de levure. Cette fois-ci, le domaine d'activation a été placé en extrémité NH₂-terminale de cette famille de protéines de fusion, afin de faciliter la construction des protéines hybrides.

Afin d'éviter un trop grand nombre de clones faux-positifs, il est préférable d'insérer la banque dans le plasmide portant le domaine d'activation, puisque Ma et Ptashne (1987b) avaient montré auparavant, que de nombreux fragments protéiques fusionnés aux 147 premiers acides aminés de Gal4 étaient capables d'activer la transcription.

Depuis l'article de Chien *et al.* (1991) les souches et plasmides utilisés ont évolué. Notamment la combinaison d'une sélection métabolique associée à un crible *lacZ* s'est généralisée. Tandis que Chien *et al.* (1991) travaillaient uniquement sur la base du crible *lacZ*, des approches plus récentes utilisent également l'acquisition de la capacité à pousser sur un milieu manquant soit d'histidine, soit de leucine. En principe seules les cellules portant une protéine hybride issue de la banque, capable d'activer l'expression du gène de sélection (His3, Leu2) via une interaction X,Y formeront des colonies sur boîte. Une telle sélection permet l'analyse d'un plus grand nombre de clones.

Parallèlement à Gal4₁₋₁₄₇ comme support de fixation à l'ADN, l'équipe de Roger Brent (Harvard Medical School) a introduit le répresseur LexA d'*Escherichia coli* comme support (Golemis et Brent, 1992 ; Gyuris *et al.*, 1993).

LexA est une protéine de 202 acides aminés, composée de deux domaines reliés par une région charnière d'environ 30 acides aminés. Les 70 premiers acides aminés constituent le domaine de fixation à l'ADN (Hurstel *et al.*, 1986; Fogh *et al.*, 1994) et les acides aminés 100 à 202 forment le domaine de dimérisation de la protéine (Schnarr *et al.*, 1985). LexA reconnaît des sites

palindromiques sur l'ADN (les opérateurs SOS) et plusieurs opérateurs adjacents peuvent être occupés d'une façon coopérative (Llobès *et al.*, 1991).

L'utilisation de LexA comme support de fixation à l'ADN peut avoir plusieurs avantages. LexA appartient à un organisme hétérologue (*Escherichia coli*), n'affecte pas la croissance de la levure, ne possède pas d'activité transcriptionnelle résiduelle (comme le segment 94-106 de Gal4) et peut être utilisé dans des souches de levure *gal+*. De ce fait, une des protéines hybrides (généralement celle contenant le domaine d'activation) peut être placée sous le contrôle de Gal4. Ainsi, la protéine hybride sera exprimée uniquement en présence de galactose, mais pas en présence de glucose. Le domaine d'activation utilisé par l'équipe de Roger Brent est également issu d'*Escherichia coli*. Il s'agit d'un des domaines activateurs artificiels isolés par Ma et Ptashne (1987b). Ce domaine (nommé B42) possède un potentiel d'activation plus faible que les domaines d'activation de Gal4 ou de VP16, ce qui permet d'éviter les problèmes de toxicité liés à un éventuel "squenching" (la titration d'un élément essentiel de la machinerie de transcription par un domaine d'activation). Une description récente du système double-hybride s'appuyant sur l'utilisation de LexA se trouve dans Golemis *et al.* (1996) et Golemis et Brent (1996).

Malgré la fréquence d'utilisation des différents systèmes double-hybride, peu d'études ont été menées cherchant à corréler des données obtenues *in vivo* avec des mesures d'affinité *in vitro*. Estojak *et al.* (1995) ont cherché à établir une telle corrélation dans le cas de plusieurs protéines à hélice-loop-hélice (Myc, Max, Mxi) et dans le cas de mutants du répresseur lambda. Leurs résultats sont quelque peu décevants.

En effet, le système double-hybride montre une forte polarité pour le couple Myc/Max. Lorsque Myc est fixé sur DBD et Max sur AD, aucune activation de la transcription n'est observée. Par contre, lorsque Max est fixé sur DBD et Myc sur AD, on observe une activation en accord avec des données biochimiques montrant que les deux protéines interagissent. Lorsque cette dernière géométrie est utilisée avec des reporters *lacZ* ayant un nombre variable d'opérateurs en amont du gène, la hiérarchie des affinités apparentes *in vivo* entre les couples Max/Myc et Max/Mxi n'est pas la même selon le reporter utilisé.

La situation est encore plus confuse dans le cas des mutants du répresseur lambda. Des mutants ayant une capacité de dimérisation fortement amoindrie activent certains gènes reporters mieux que la protéine sauvage (Estojak *et al.*, 1995). Par ailleurs, bien que tous les variants activent assez fortement le gène *lacZ*, trois sur quatre sont incapables d'activer le gène *leu2* et auraient, de ce

fait, échappé à une sélection sur milieu dépourvu de leucine. En conclusion, Estojak *et al.* (1995) estiment qu'aucun gène reporter montre une corrélation satisfaisante avec les affinités mesurées *in vitro*.

Etant donné la complexité de l'assemblage d'un complexe d'initiation de la transcription chez les eucaryotes (voir : Hori et Carey, 1994 ; Maldonado et Reinberg, 1995) qui mobilise au moins une trentaine de protéines différentes, et le fonctionnement toujours mal compris des activateurs de la transcription, ces résultats ne sont pas tellement surprenants.

LexA est le répresseur du système SOS chez *E.coli*. LexA, composé de deux domaines, est inactivé *in vivo* suite à des dommages sur l'ADN par un clivage protéolytique localisé entre les deux domaines (pour une revue sur LexA, voir Schnarr et Granger-Schnarr, 1993). La séparation du domaine de fixation à l'ADN du domaine de dimérisation entraîne alors une diminution de l'affinité pour l'ADN d'environ 1000-fois, ne permettant plus une répression transcriptionnelle efficace. D'autre part, la constante de dimérisation de LexA ($K_A = 2 \times 10^4 \text{ M}^{-1}$) est relativement faible (Schnarr *et al.*, 1985), ce qui suggère que la greffe d'un domaine de dimérisation hétérologue à affinité modeste devrait reconstituer un répresseur hybride fonctionnel. Schmidt-Dörr *et al.* (1991) ont pu montrer en effet que la fusion de la glissière à leucine ("leucine-zipper") de l'oncoprotéine Jun à LexA₁₋₈₇ restaure une répression transcriptionnelle. Par contre, la fusion de la glissière à leucine Fos ("leucine-zipper Fos"), qui est incapable de conférer l'homodimérisation, ne restaure pas la répression (Porte *et al.*, 1995). En partant de cette observation, la partie glissière de Fos ("Fos-zipper") de la protéine de fusion a ensuite été mutagénisée, afin de caractériser les anti-déterminants de l'homodimérisation de Fos (Porte *et al.*, 1995).

La souche bactérienne indicatrice est déficiente en LexA et porte le gène *lacZ* sous le contrôle d'un opérateur SOS reconnu par LexA (l'opérateur *sulA*). La protéine hybride, placée sous le contrôle du promoteur *lacUV5* inductible par l'IPTG, est produite à partir d'un plasmide de façon à pouvoir adapter la quantité de protéine aux nécessités expérimentales. Ce système est parfaitement adapté à l'étude d'un complexe homodimérique, mais son utilisation devient laborieuse du moment où l'on veut étudier l'hétérodimérisation entre deux protéines différentes capables de former également des homodimères.

Parallèlement au système LexA, l'équipe de James Hu a développé une approche voisine en utilisant le domaine de fixation à l'ADN du répresseur lambda (Hu, 1995). L'avantage de ce système est d'offrir les possibilités de sélection du phage lambda. Son inconvénient réside dans le fait que le domaine

de fixation à l'ADN du répresseur lambda contient déjà un élément de dimérisation (l'hélice 5), qui peut masquer l'effet de domaines de dimérisation hétérologues relativement faibles.

L'invention a notamment pour objet de résoudre les difficultés présentées par les systèmes de l'art antérieur et, en particulier, l'invention a pour but de permettre l'étude de l'hétérodimérisation.

L'invention a pour objet un système permettant la détection d'interaction entre deux protéines greffées, par visualisation de la répression de la transcription et non pas par une activation de la transcription. En effet, contrairement au système levure, une protéine ayant une activité activatrice intrinsèque pourrait être utilisée.

L'invention a pour objet un système de détection des interactions protéine-protéine à l'aide d'une bactérie (*Escherichia coli*) comme organisme hôte, couramment utilisée dans tous les laboratoires de biologie moléculaire.

L'invention a pour objet un système de détection des interactions protéine-protéine évitant le problème de l'adressage des protéines de fusion dans le noyau comme c'est le cas dans la levure et ce, grâce à la non-compartmentation d'*E. coli*, notamment à l'absence d'un noyau structuré.

L'invention a pour objet un système adapté à l'étude des interactions de protéines situées dans la membrane cytoplasmique.

L'invention a pour objet un test (répression de la transcription) permettant potentiellement l'identification de protéines interagissant avec les domaines d'activation des facteurs de transcription, parmi lesquels on trouve de nombreux produits oncogéniques impliqués dans la cancérogénèse, comme Jun, Fos ou les membres de la famille Ets.

L'invention a pour objet un système d'étude d'interaction protéine-protéine (détermination des forces d'interaction, identification des éléments de structure ou de séquence importants pour cette interaction, analyse de la spécificité de l'interaction) plus fiable que le système levure (Estojak *et al.* 1995).

L'invention a pour objet un système d'étude d'interaction protéine-protéine permettant de détecter des interactions lorsque la protéine "hameçon" (celle pour laquelle on recherche le ou les partenaires) a des capacités intrinsèques d'activation de la transcription (revues Bartel *et al.* 1993, Golemis et Brent 1996, Golemis *et al.* 1996) et ceci se produisant assez fréquemment (Ma et Ptashne 1988), et donc l'invention a pour objet un système permettant l'étude des interactions d'un "vrai" domaine d'activation de la transcription ;

L'invention a pour objet un système de détection d'interaction protéine-protéine permettant l'examen d'un grand nombre de clones ($>10^6$) grâce à l'efficacité de transfection de *E. Coli* qui est nettement supérieure à celle de la levure (de l'ordre de 1 000 fois) et ainsi le criblage de banques d'ADNc ou d'ADN génomique.

L'invention concerne l'utilisation d'une séquence d'ADN comprenant :

- une séquence nucléotidique codant pour une protéine indicatrice (gène indicateur),

- un promoteur,

- un opérateur muté,

ladite séquence nucléotidique étant sous le contrôle du susdit promoteur comprenant le susdit opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte une mutation 2, les homodimères susceptibles d'être formés à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

- pour détecter ou quantifier des interactions entre deux protéines ou domaines protéiques, ou cribler des protéines ou domaines protéiques présentant des interactions avec une protéine ou domaine protéique déterminé(e), ces interactions étant révélées lorsqu'il y a répression de la transcription du susdit gène indicateur, cette répression résultant de la formation d'hétérodimères comprenant :

- 1) un premier monomère contenant une première protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence de l'opérateur muté contenant la mutation 1, le susdit site de fixation étant fusionné à l'une des deux protéines ou à l'un des deux domaines protéiques, impliqué(e) dans les susdites interactions,

2) un deuxième monomère contenant une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon silencieuse ou comportant la mutation 2 de l'opérateur muté, le susdit site de fixation étant fusionné à l'autre des deux protéines ou à l'autre des deux domaines protéiques, impliqué(e) dans les susdites interactions.

L'utilisation selon l'invention repose sur le fait qu'une protéine régulatrice, par exemple LexA, régule *in vivo* l'expression de gènes en se fixant sur des séquences spécifiques situées en amont de ces gènes (séquences opérateur).

L'opérateur utilisé dans le cadre de l'invention est un opérateur muté dérivant d'un opérateur non muté qui est tel que, à l'état sauvage, ses séquences sont palindromiques (répétées inversées) et reconnues par des homodimères de la protéine régulatrice. La protéine régulatrice, par exemple LexA, est organisée en deux domaines, un domaine de fixation à l'ADN qui reconnaît spécifiquement les séquences opérateur et un domaine de dimérisation par lequel se font les interactions entre deux monomères protéiques stabilisant l'interaction des dimères sur l'ADN. Chaque monomère se fixe sur un demi-opérateur. Par "demi-opérateur" on désigne la séquence spécifiquement reconnue par chaque monomère. Ces séquences spécifiques sont identiques dans le cadre d'un opérateur sauvage ou doublement muté, et différentes dans le cadre d'un opérateur muté.

Cette fixation est renforcée par l'existence de contacts entre les domaines de dimérisation de chaque monomère. Le domaine de fixation à l'ADN seul (dépourvu du domaine de dimérisation) est incapable de fixer efficacement l'opérateur et donc de réguler l'expression d'un gène. Si on remplace le domaine de dimérisation de LexA par des motifs protéiques hétérologues mais susceptibles d'interagir, on restitue l'activité régulatrice de LexA.

Ceci étant, dans le cadre de l'invention, la séquence opérateur reconnue par une protéine régulatrice, par exemple LexA, a été modifiée de telle sorte qu'un demi-opérateur restant sauvage sera reconnu par la protéine régulatrice sauvage, par exemple LexA, tandis que l'autre demi-opérateur sera muté et reconnu par une forme mutée de la protéine régulatrice.

Chacun de ces domaines de fixation à l'ADN peut être fusionné à des domaines protéiques susceptibles d'hétérodimériser et toute interaction entre ces domaines est mesurée par une répression plus ou moins importante du gène indicateur (également désigné par "gène reporter").

On entend par gène indicateur :

- soit un gène codant pour une protéine dont l'activité d'une part est révélée par la couleur qu'elle confère aux clones bactériens la produisant et d'autre part peut être dosée quantitativement, par exemple le gène *lacZ*, le gène codant pour la luciférase,

- soit un gène codant pour une protéine toxique pour la bactérie de telle sorte que si la protéine est produite, la bactérie meure, par exemple le gène *sulA*.

On peut également utiliser un système pour lequel :

- le gène toxique est un gène codant pour une résistance à un antibiotique,
- l'opérateur mixte *op408/opWT* est placé dans le promoteur du gène codant pour le répresseur de ce gène de résistance à un antibiotique.

En absence d'interaction entre *LexA408-X* et *LexAWT-Y*, le répresseur du gène en question est produit, la bactérie ne peut pas pousser sur un milieu de culture contenant cet antibiotique. Dans le cas contraire, la synthèse du répresseur n'a pas lieu, le gène de résistance à l'antibiotique est produit et les bactéries donnent lieu à des colonies sur un milieu de culture contenant cet antibiotique.

Par "homodimères", on désigne des associations non covalentes de protéines dans lesquelles les domaines d'interaction protéine-protéine sont identiques et liés de façon covalente à des domaines protéiques hétérologues ou homologues de fixation de l'ADN, eux-mêmes identiques ou différents.

Par "hétérodimères", on désigne des associations non covalentes de protéines dans lesquelles les domaines d'interaction protéine-protéine sont différents et liés de façon covalente à des domaines protéiques hétérologues ou homologues de fixation de l'ADN, eux-mêmes identiques ou différents.

Par "mutation silencieuse portée par l'une des séquences de l'opérateur", on désigne une mutation qui n'affecte pas la reconnaissance de ladite séquence par la protéine régulatrice sauvage.

Par "mutation 2 portée par la séquence de l'opérateur", on désigne une mutation différente de la mutation 1, et qui est telle que la séquence comportant ladite mutation 2 est reconnue par une protéine régulatrice non susceptible de reconnaître la séquence portant la mutation 1, la mutation 1 étant elle-même telle qu'elle est reconnue par une protéine régulatrice qui n'est pas susceptible de reconnaître la séquence portant la mutation 2.

Par "interaction protéine-protéine", on désigne l'assemblage non covalent de deux (ou plusieurs) protéines par interaction Van der Waals (interactions électrostatiques entre charges opposées, liaisons hydrogène, interactions

hydrophobes). La surface masquée par une telle interaction est généralement de 1400 à 1600 Å² (pour une revue voir: J. Janin "Principles of protein/protein recognition from structure to thermodynamics" 1995, Biochimie 77:497-505).

Par "domaine de fixation à l'ADN d'une protéine", on définit la partie de la protéine dont l'intégrité est indispensable à sa fixation non covalente à l'ADN.

Par "détection de l'interaction", on désigne une méthode permettant de visualiser une interaction entre deux protéines.

Par "quantification de l'interaction", on désigne une méthode permettant de chiffrer la force de cette interaction.

L'expression "protéine susceptible de reconnaître la séquence d'un opérateur" désigne une protéine capable de faire des liaisons non covalentes avec une séquence particulière d'ADN.

L'expression "domaine de fixation fusionné à une protéine" désigne un domaine protéique relié par une liaison peptidique à une autre protéine ou domaine de protéine.

Selon un mode de réalisation avantageux de l'invention, la protéine régulatrice est dépourvue de son domaine de dimérisation, lequel peut toutefois être toléré si les interactions résultant de la présence du domaine de dimérisation sont plus faibles que celles que l'on veut mettre en évidence.

L'invention a également pour objet une séquence d'ADN comprenant :

- une séquence nucléotidique codant pour une protéine indicatrice (gène indicateur),
- un promoteur,
- un opérateur muté,

ladite séquence nucléotidique étant sous le contrôle d'un promoteur comprenant un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence

qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte une mutation 2, les homodimères susceptibles d'être formés à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

ladite mutation 1 étant en outre telle qu'elle est reconnue spécifiquement par une protéine régulatrice comportant un site de fixation à l'ADN entrant dans la constitution d'une première protéine de fusion, lequel site de fixation est fusionné avec une protéine ou un domaine protéique susceptible d'interagir avec une autre protéine ou un autre domaine protéique entrant dans la constitution d'une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon silencieuse ou comportant la mutation 2.

L'invention concerne une séquence d'ADN caractérisée en ce qu'elle comprend un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), l'une de ces deux séquences étant mutée, et l'autre étant à l'état sauvage.

La séquence mutée peut être en amont ou en aval de la séquence à l'état sauvage.

L'invention a également pour objet une séquence d'ADN caractérisée en ce que l'opérateur muté contient les séquences suivantes :

CCGT et ACAG

espacées par une séquence d'environ 3 à environ 20 nucléotides, et notamment de 8 à 12, et avantageusement 8 nucléotides, et notamment caractérisée en ce que l'opérateur muté contient la séquence suivante :

5'-CGGAATTC AATAGGGTTGATCTTTGTTGTC ACTGGATGTACCGTACATCCATACAGTAACTCACAGGGGC-3'

Les séquences CCGT et ACAG correspondent aux séquences respectivement CTGT et ACAG de l'opérateur non muté.

La présente invention concerne une séquence d'ADN caractérisée en ce que

la séquence nucléotidique codant pour la protéine indicatrice est le gène *lacZ* chez *E. coli*, codant pour la β -galactosidase,

l'opérateur muté contient deux séquences et dérive d'un opérateur contenant deux séquences répétées inversées (palindrome), l'une de ces deux séquences

étant mutée, et l'autre étant à l'état sauvage, la séquence mutée étant de préférence celle située en aval de la séquence à l'état sauvage,

la séquence sauvage étant reconnue spécifiquement par la protéine LexA sauvage ou par un fragment de LexA contenant le site de fixation à l'ADN notamment un fragment défini par les positions 1 à x de LexA, x variant de 69 à 87, et notamment par les fragments 1 à 69 ou 1 à 87 de LexA,

et la séquence mutée étant reconnue par la protéine LexA mutée ou par un fragment muté de LexA contenant le site de fixation à l'ADN, notamment un fragment muté de LexA correspondant à celui défini par les positions 1 à x de LexA, x variant de 69 à 87, et notamment par les fragments 1 à 69 ou 1 à 87, et les mutations de LexA étant avantageusement $P40 \rightarrow A$, $N41 \rightarrow S$, $A42 \rightarrow S$, le domaine de dimérisation de LexA étant absent à la fois de LexA ou du fragment de LexA et de LexA muté ou du fragment muté de LexA.

Les trois mutations simultanées $P40 \rightarrow A$, $N41 \rightarrow S$, $A42 \rightarrow S$ dans LexA donnent lieu à un mutant désigné par LexA408.

Par "domaine de dimérisation", on désigne la partie de la protéine dont l'intégrité est indispensable à l'interaction non covalente avec elle-même (homodimère) ou avec un autre domaine de dimérisation (hétérodimère).

Il est nécessaire que le domaine de dimérisation de LexA et de LexA408 ou du fragment utilisé de LexA et de LexA408 soient absents sinon lorsque les deux protéines (LexA et LexA408) sont coexprimées dans l'hôte cellulaire impliqué dans l'invention, le gène indicateur serait réprimé.

La présente invention a également pour objet un complexe entre

- une séquence d'ADN comprenant une séquence nucléotidique codant pour une protéine indicatrice, ladite séquence nucléotidique étant sous le contrôle d'un promoteur comprenant un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte

une mutation 2, les homodimères susceptibles d'être fournis à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

- et des hétérodimères dont les monomères les constituant (respectivement premier monomère et deuxième monomère) sont impliqués entre eux dans une interaction protéine/protéine :

1) le premier monomère étant tel qu'il contient une première protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence de l'opérateur muté contenant la mutation 1, le susdit site de fixation étant fusionné à l'une des deux protéines ou à l'un des deux domaines protéiques, impliqué(e) dans les susdites interactions,

2) le deuxième monomère étant tel qu'il contient une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon silencieuse ou comportant la mutation 2 de l'opérateur muté, le susdit site de fixation étant fusionné à l'autre des deux protéines ou à l'autre des deux domaines protéiques, impliqué(e) dans les susdites interactions.

La présente invention concerne également un complexe caractérisé en ce que :

- la séquence d'ADN est telle que définie ci-dessus
et en ce que

- les hétérodimères contiennent :

1) un monomère contenant une première protéine de fusion contenant le site de fixation à l'ADN de LexA muté ou d'un fragment muté de LexA reconnaissant la séquence de l'opérateur muté contenant la mutation 1, lequel site de fixation est fusionné avec une protéine déterminée ou un fragment protéique déterminé susceptible d'être impliqué dans une interaction tel que décrite ci-dessus, laquelle protéine ou lequel domaine protéique contient par exemple un domaine d'activation ou de dimérisation,

2) un monomère contenant une deuxième protéine de fusion contenant le site de fixation à l'ADN de LexA sauvage, ou porteur d'une mutation silencieuse ou d'une mutation 2 ou d'un fragment de LexA reconnaissant la séquence non mutée de LexA ou mutée de façon silencieuse ou comportant une mutation 2, lequel site de fixation est fusionné avec une protéine déterminée ou un fragment protéique déterminé, contenant par exemple une protéine ou un domaine protéique susceptible d'interagir avec la protéine ou le fragment

protéique défini en 1) ou avec les susdits domaines d'activation ou de dimérisation.

L'invention a également pour objet un hôte cellulaire contenant une séquence d'ADN telle que décrite ci-dessus.

L'invention a également pour objet un procédé pour détecter et quantifier des interactions entre deux protéines ou fragments de protéines, caractérisé

- en ce que l'on met en présence des séquences nucléotidiques codant respectivement pour le premier et le deuxième monomère contenant respectivement une première et une deuxième protéine de fusion telle que définie ci-dessus avec une séquence d'ADN telle que décrite précédemment, et

- en ce que l'on détecte l'éventuelle répression de la transcription du gène indicateur tel que défini ci-dessus, notamment par transfection d'un hôte cellulaire tel que décrit ci-dessus, avec des vecteurs contenant, d'une part, les susdites séquences nucléotidiques codant respectivement pour le premier et le deuxième monomère contenant respectivement la première et la deuxième protéine de fusion et, d'autre part, les éléments nécessaires à l'expression des susdits monomères.

L'invention concerne un procédé pour identifier une protéine interagissant avec une protéine déterminée, caractérisé

- en ce que l'on met en présence les éléments suivants :

- . une séquence nucléotidique codant pour un premier monomère contenant une première protéine de fusion entre, d'une part, une protéine régulatrice selon l'invention et, d'autre part, la protéine déterminée,

- . une banque génomique, contenant notamment des fragments d'ADN, et notamment d'ADNc, chacun étant susceptible de coder pour une protéine interagissant avec la protéine déterminée, chacun des différents éléments de la banque génomique étant respectivement fusionné à la séquence codant pour une protéine régulatrice telle que définie ci-dessus,

- . une séquence d'ADN selon l'invention,

et

- en ce que l'on détecte l'éventuelle répression de la transcription de la séquence nucléotidique codant pour une protéine indicatrice telle que décrite ci-dessus, notamment par transfection d'un hôte cellulaire selon la présente invention, d'une part, avec un vecteur contenant la séquence nucléotidique codant pour le susdit premier monomère contenant la première protéine de fusion et, d'autre part, un ensemble de vecteurs contenant une séquence nucléotidique codant pour un deuxième monomère contenant une deuxième protéine de fusion, chaque séquence nucléotidique codant pour chaque deuxième

monomère contenant la séquence nucléotidique codant pour la protéine régulatrice selon la présente invention, et en aval de celle-ci l'un des éléments d'ADN de la susdite banque, tous les vecteurs contenant en outre les éléments nécessaires à l'expression des susdits monomères.

Ce procédé fait l'objet de l'exemple ci-après. Dans cet exemple, le gène indicateur est *lacZ* et a été placé sous le contrôle d'une séquence promoteur/opérateur caractérisée en ce que l'opérateur a été muté comme indiqué ci-après. Ceci a donné lieu à la construction d'une souche indicatrice (SU202). Cette souche permet d'identifier, sur la base de la couleur des clones bactériens, l'existence d'interactions protéine-protéine entre deux protéines ou domaines protéiques fusionnés respectivement au domaine de fixation à l'ADN de LexA sauvage et LexA408 tels que décrits ci-après. Cette souche permet de quantifier et d'analyser par mutagenèse cette interaction.

Cependant, le criblage d'une banque d'ADN nécessite l'analyse au minimum de 10^6 clones. Ceci implique l'obtention sur milieu solide indicateur approprié (comme indiqué sur la figure 5) et l'expérience de 10^6 colonies bactériennes dont il faudrait vérifier la couleur (blanches ou roses en cas d'interaction et rouges en absence d'interaction). Bien que théoriquement faisable, ceci est très coûteux non seulement en temps, mais également en matériel jetable.

Pour identifier de nouvelles interactions, on a avantageusement recours au système de sélection suivant. Dans ce type d'application, les interactions protéine-protéine ne sont plus visualisées sur la base de la couleur des clones bactériens, mais sur le simple fait de leur existence. En effet, si le gène indicateur est remplacé par un gène codant pour une protéine toxique :

- l'absence d'interaction protéine-protéine entre les parties Y et X de LexA₁₋₈₇WT-Y et de LexA₁₋₈₇408-X (comme décrit dans la figure 1) entraîne la production de la protéine toxique et donc la mort de la bactérie,
- l'existence de ces interactions entraîne la répression du gène toxique, la bactérie peut alors se diviser et donner lieu à une colonie sur milieu de culture solide.

Un tel système basé sur la survie des bactéries est bien évidemment plus facile à mettre en oeuvre. Il est cependant évident que, comme dans tous les systèmes de sélection, il faut compter avec ce que l'on appelle des "faux positifs", c'est-à-dire des clones qui survivent pour des raisons autres que celles que l'on identifie, par exemple, des mutations spontanées qui inactivent le produit du gène toxique. Pour lever cette ambiguïté, on utilise la souche indicatrice SU202. En effet, seuls les clones poussant blancs ou roses sont

considérés comme "vrai positifs", les clones survivants dus à une mutation dans le gène toxique étant incapables de réprimer *LacZ* et poussant rouges.

Le gène toxique utilisé est le gène *sulA*. Il code pour une protéine, SULA, inhibitrice de la division cellulaire. En sa présence, la bactérie ne se divise plus et meurt.

Pour l'instant, la construction consistant à mettre *sulA* sous le contrôle d'un opérateur muté a été menée à bien. En utilisant le système modèle décrit ci-après avec les protéines hybrides de LexA₁₋₈₇408-cJunZip et LexA₁₋₈₇WT-cFosZip, on a montré que le système fonctionne. En effet, la construction (*sulA* sous le contrôle d'un opérateur muté) ne peut être propagée dans un hôte bactérien que si celui-ci porte également les gènes codant pour les protéines hybrides de LexA₁₋₈₇408-cJunZip et LexA₁₋₈₇WT-cFosZip.

La mise en évidence de nouvelles interactions protéine-protéine nécessite la constitution d'une banque d'ADN liguée en fusion à la suite du domaine de fixation. Pour cela, on a constitué une banque d'ADN qui est liguée en fusion à la suite du domaine de fixation à l'ADN de LexA sauvage ou d'un mutant de LexA reconnaissant la partie sauvage de l'opérateur mais avec une affinité moindre tels que ceux portant une des mutations suivantes G → D ou G → S en position 23 (réf. 21). Cette alternative a pour but de limiter l'apparition de clones "faux positifs" qui peuvent être dus à une interaction homodimérique trop forte d'une protéine ou domaine protéique présent dans la banque.

Description des figures :

- **Figure 1** : la Figure 1 concerne le principe du fonctionnement du système double-hybride chez *E. coli*

Le système est basé sur le fait qu'un opérateur muté contenant 1/2 opérateur sauvage (WT) et 1/2 opérateur muté (408) n'est reconnu que par un dimère mixte LexAWT/LexA408. Un gène indicateur (*lacZ* représenté par un rectangle vide) a été mis sous le contrôle de cet opérateur mixte (représenté par un rectangle noir pour la partie non mutée et annotée WT et par un rectangle hachuré pour la partie mutée et annotée 408). Des protéines ou des sous domaines de protéines X et Y susceptibles d'interagir sont fusionnés aux domaines de fixation à l'ADN de LexA₁₋₈₇WT (représenté par une sphère noire) et de LexA₁₋₈₇408 (représentée par une sphère hachurée). LexA₁₋₈₇WT-X (représenté par un rectangle hachuré relié à une sphère hachurée) et LexA₁₋₈₇408-Y (représenté par un rectangle noir relié à une sphère noire) sont clonés sur des plasmides compatibles et leur expression est contrôlée par un promoteur inductible (*lacUV5*, Oertel-Buchheit P., Lamerichs R.M.J.N.,

Schnarr M. and Granger-Schnarr M. 1990 Mol.Gen.Genet. 223:40-48 *Genetic analysis of the LexA repressor: isolation and characterization of LexA(Def) mutant proteins.*). Les plasmides sont représentés par des cercles dont une partie épaissie est séparée en deux, l'une représente le gène codant pour les protéines, LexA₁₋₈₇WT-X ou LexA₁₋₈₇₄₀₈-Y et l'autre pour leur séquence promoteur, lacUV5). Ces protéines, LexA₁₋₈₇WT-X et LexA₁₋₈₇₄₀₈-Y peuvent exister soit sous forme libre dans la bactérie, soit sous forme liée à l'ADN (représentée par un rectangle noir relié à une sphère grisée entrelacé avec un rectangle hachuré relié à une sphère hachurée eux mêmes reliés aux rectangles noirs et hachurés représentant l'opérateur mixte ci-dessus décrit.). Le promoteur LacUV5 est lui même sous le contrôle d'une protéine régulatrice, le répresseur lactose (représenté par un carré aux coins arrondis et segmenté en cartiers). Le gène codant pour ce répresseur (gène *lacI*) est cloné sur un épisome, le facteur F' (Willetts N. and Skurray R. in *E. coli and Salmonella Typhimurium, Cellular and Molecular Biology* Ed. Neidhardt F.C. 1987, pp 1110-1133 et Oertel-Buccheit *et al* 1990) lui même représenté par un arc de cercle. Le répresseur lactose peut exister soit sous forme libre dans la bactérie, soit sous forme liée à l'ADN (représenté par un carré aux coins arrondis et segmenté en cartiers relié aux séquences promoteurs, lacUV5 décrites ci-dessus). Le terme LexA(Def) signifie que la souche est dépourvue de LexA endogène.

- **Figure 2** : la Figure 2 représente la construction de la souche indicatrice pour la mise en oeuvre d'un système double hybride chez *E. coli*.

Le plasmide pRS415 (représenté par le cercle contenant à l'intérieur l'indication "pRS415" (SIMONS R.W., HOUMAN F. & KLECKNER N. (1987). *Improved single and multicopy lac-based cloning vectors for protein and operon fusions*. *Gene* 53:85-96) porte le gène *lacZ*, dépourvu de région promoteur/opérateur (représenté par un épaississement du trait) et deux sites uniques de coupure pour deux enzymes de restriction EcoRI et BamHI. "Lac⁻" signifie que ce plasmide est incapable de produire le produit du gène *lacZ* puisqu'il est dépourvu d'une séquence promoteur/opérateur. Ce plasmide est coupé par ces deux enzymes puis ligué à un oligonucléotide de 70 paires de bases (70mer et représenté par un rectangle noir) dont la séquence est la suivante :

GAATTCAATAGGGTTGATCTTTGTTGTCACTGGATGTACCGTACA
TCCATACAGTAACTCACAGGGGCACCCTAGG-3' cet oligonucléotide introduit ainsi devant le gène *lacZ* une séquence promoteur/opérateur permettant au gène *lacZ* d'être transcript et régulé par l'opérateur mixte tel qu'il est défini dans la légende de la figure 1. Cette ligation symbolisée par une flèche entre le

plasmide et le 70mer crée le plasmide pGC202 (représenté par le cercle contenant à l'intérieur l'indication "pGC202"). "Lac⁺" signifie que ce plasmide est maintenant capable de produire le produit du gène *lacZ* puisqu'il est pourvu d'une séquence promoteur/opérateur. Le gène indicateur ainsi constitué (représenté par un rectangle noir pour la séquence promoteur/opérateur et par un rectangle blanc pour le gène *lacZ*) est ensuite transféré sur l'ADN d'un phage lambda (λ) par un double événement de recombinaison entre l'ADN plasmidique et phagique (on entend par double événement de recombinaison un mécanisme moléculaire complexe se produisant *in vivo* et aboutissant à l'échange de fragments d'ADN entre des molécules partiellement homologues, ceci est matérialisé sur la figure par deux croix entre le plasmide pGC 202 et une droite symbolisant l'ADN du phage) donnant lieu au phage λ GC 202. Ce phage a la capacité de s'intégrer dans le chromosome bactérien par un mécanisme complexe ayant lieu *in vivo* (Elledge *et al.* 1991). L'intérêt de ce système réside dans le fait que le gène indicateur sera sur le chromosome bactérien et non plus sur un plasmide, ce qui permet d'utiliser les vecteurs plasmidiques comme véhicule des protéines de fusion.

- **Figure 3** : la Figure 3 concerne la caractérisation de la souche SU202 en fonction de la quantité de protéines présentes dans la cellule.

En abscisse, figurent les concentrations en IPTG (M) et en ordonnées, l'expression relative de la β -galactosidase.

La variation de la quantité de protéines est obtenue en cultivant les bactéries en présence de concentrations variables en IPTG, les gènes codants pour les protéines chimères étant sous le contrôle du promoteur inductible lacUV5. Les unités β -galactosidases sont mesurées selon la méthode décrite par Miller (1972). L'expression relative de la β -galactosidase est calculée en faisant le rapport entre les unités β -galactosidases mesurées en présence des vecteurs d'expression et celles mesurées en leur absence, ces dernières correspondant à l'état non réprimé du gène *lacZ*. Les carrés blancs correspondent aux valeurs mesurées lorsque la souche contient un plasmide exprimant LexA₁-87408-cJunZip (pDL606, voir figures 6 et 7). Les cercles blancs correspondent aux valeurs mesurées lorsque la souche contient un plasmide exprimant le domaine de fixation à l'ADN seul (JWL96, Little JW and Hill S.A. (1985) *Deletion within the hinge region of a specific DNA-binding protein* Proc. Natl. Acad. Sci USA, 82:2301-2305). Les triangles blancs correspondent aux valeurs mesurées lorsque la souche contient un plasmide exprimant LexA₁-87WT-cFosZip (pMS404, voir figures 8 et 9). Les carrés noirs correspondent aux valeurs mesurées lorsque la souche contient à la fois le plasmide exprimant le domaine

de fixation à l'ADN seul (JWL96) et celui exprimant LexA₁-87408-cJunZip (pDL606, voir figures 6 et 7). Les cercles noirs correspondent aux valeurs mesurées lorsque la souche contient à la fois le plasmide exprimant LexA₁-87408-cJunZip (pDL606, voir figures 6 et 7) et celui exprimant LexA₁-87WT-cFosZip (pMS404, voir figures 8 et 9).

- **Figure 4** : la Figure 4 représente la souche bactérienne SU202 qui permet de mesurer sélectivement l'hétérodimérisation de protéines de fusion susceptibles d'homodimériser.

En abscisse, figurent les concentrations en IPTG (M) et en ordonnées, l'expression relative de la β -galactosidase.

Des mutants de la glissière à leucine ("leucine zipper") de Fos ont été construits au laboratoire, ces mutants ont acquis la capacité à homodimériser (contrairement au leucine zipper de Fos qui lui en est incapable). Quatre de ces mutants sont utilisés ici et sont décrits dans: Porte D., Oertel-Buccheit P., Granger-Schnarr M. and Schnarr M. 1995 J. Biol. Chem. **270**:22721-22730 *Fos leucine zipper variants with increased association capacity*, Il s'agit :

- * d'un mutant pour lequel les cinq positions "a" de la glissière ("zipper") (définies dans Porte *et al.* 1995) sont occupées par des résidus leucine (LexA₁-87WT-cFosZipLLLLL) et qui est représenté par des cercles blancs dans les figures 4A, B et C,

- * d'un mutant pour lequel les cinq positions "a" de la glissière ("zipper") (définies dans Porte *et al.* 1995) sont occupées successivement par deux valines, une isoleucine, une phénylalanine et une valine (LexA₁-87WT-cFosZipVVIFV) et qui est représenté par des cercles noirs dans les figures 4A, B et C,

- * d'un mutant pour lequel les cinq positions "a" de la glissière ("zipper") (définies dans Porte *et al.* 1995) sont occupées successivement par deux isoleucines, une leucine et deux isoleucines (LexA₁-87WT-cFosZipIILII) et qui est représenté par des carrés blancs dans les figures 4A, B et C,

- * d'un mutant pour lequel les cinq positions "a" de la glissière ("zipper") (définies dans Porte *et al.* 1995) sont occupées successivement par une isoleucine, une leucine, une phénylalanine, une leucine et une phénylalanine (LexA₁-87WT-cFosZipILFLF) et qui est représenté par des carrés noirs dans les figures 4A, B et C,

- * les triangles noirs représentent LexA₁-87WT-cFosZip sauvage,

- * les courbes en pointillés représentent les mesures faites en absence de plasmide.

Chacune de ces constructions est soit introduite séparément dans les souches bactériennes indicatrices (figure 4A pour la souche SU202 et 4B pour la souche SU101) soit co-introduites avec la protéine de fusion LexA₁-87408-cJunZip (figure 4C). L'activité β -galactosidase est ensuite mesurée dans les conditions décrites par Miller *et al.* 1972 en fonction de concentrations variables en IPTG. En absence de protéine, le nombre d'unités mesurées reste constant, indiquant que le gène codant pour la β -galactosidase est complètement exprimé, lorsque le nombre d'unités mesurées diminue, cela signifie que le gène codant pour la β -galactosidase est réprimé, l'intensité de la répression étant fonction à la fois de la quantité de protéine présente dans la cellule (elle augmente lorsque l'on augmente la concentration en IPTG) et de la force d'interaction des protéines entre elles.

A- La souche SU202 (comme décrite dans cette revendication) ne "répond" pas à l'homodimérisation (*lacZ* contrôlé par un opérateur mixte WT/408).

B- L'homodimérisation des différentes protéines est mise en évidence dans une souche (SU101) dans laquelle *lacZ* est contrôlé par un opérateur WTC-La souche SU202 mesure sélectivement leur hétérodimérisation avec LexA408JunZip

- **Figure 5** : sur la Figure 5, on voit que la souche SU 202 est transformée soit avec chacun des plasmides portant les gènes codant pour les protéines de fusion LexA₁-87408-cJunZip et LexA₁-87WT-cFosZip soit cotransformée avec les deux plasmides. Les cellules transformées sont laissées une nuit à température ambiante dans du LB additionné respectivement soit de kanamycine, chloramphénicol, ampicilline soit de kanamycine, chloramphénicol, tétracycline soit des quatre antibiotiques: kanamycine, chloramphénicol, ampicilline et tétracycline, avant d'être étalées sur milieu solide McConkey contenant les mêmes antibiotiques. Les cellules transformées par un seul des deux plasmides poussent rouges. Cette coloration signifie que le gène de la β -galactosidase n'est pas réprimé par les protéines de fusion et que l'enzyme produit fermente le lactose du milieu. Les bactéries cotransformées par les deux plasmides poussent blanches, le gène de la β -galactosidase est réprimé.

- **Figure 6** : la Figure 6 représente la carte du plasmide pDL606 (pour le clonage de "l'hameçon"). Les sites de restrictions indiqués sont des sites uniques. Ap représente la gène codant pour la résistance à l'ampicilline, p15A l'origine de répllication du plasmide. Le vecteur est de type pACYC (New England Biolabs, Inc) : pACYC177.

- **Figure 7** : la Figure 7 représente la séquence d'une partie du plasmide pDL606 codant pour la protéine de fusion LexA₁₋₈₇408-cJunZip. Les sites indiqués permettent de se repérer sur la carte de la figure 6.

- **Figure 8** : la Figure 8 représente la carte du plasmide pMS404. Les sites de restriction indiqués sont des sites uniques. Tet représente le gène codant pour la résistance à la tétracycline, Ori l'origine de réplication du plasmide. Le vecteur est de type pUC (*New England Biolabs, Inc*).

- **Figure 9** : la Figure 9 représente la séquence d'une partie du plasmide pMS404 codant pour la protéine de fusion LexA₁₋₈₇WT-cFosZip. Les sites indiqués permettent de se repérer sur la carte de la figure 8.

- **Figure 10** : la Figure 10 représente la carte du plasmide pL23B (vecteur plasmidique pour clonage de banque pour SU202). Les sites de restriction indiqués sont des sites uniques. Tet représente le gène codant pour la résistance à la tétracycline, Ori l'origine de réplication du plasmide. Ce plasmide porte le domaine de fixation à l'ADN d'un mutant de LexA, LexA₁₋₈₂3 (mutation G->S en position 23). Cette mutation n'affecte pas la spécificité de reconnaissance (LexA₂₃ reconnaît la même séquence que LexA_{WT}) mais affecte la force avec laquelle cette interaction protéine/ADN se produit. En effet, pour améliorer le système, nous avons trouvé qu'en diminuant la force d'interaction entre LexA et l'ADN, on gagnait en sensibilité. Ce plasmide est prévu pour le clonage d'une banque.

- **Figure 11** : la Figure 11 représente la séquence d'une partie du plasmide pL23B codant pour LexA₁₋₈₇23. La banque peut être clonée derrière LexA₁₋₈₇23 en utilisant l'un des sites uniques suivants: XhoI, EcoRI, SfiI. En aval de ces sites trois codons stops (TAA) ont été insérés dans les trois cadres de lecture afin de limiter la protéine de fusion. Les sites indiqués permettent de se repérer sur la carte de la figure 10.

L'invention sera complétée par la description détaillée et les exemples qui suivent, donnés à titre illustratif et non limitatif.

Principe du fonctionnement du système double-hybride chez *E. coli*

Le système double-hybride chez *E. coli* est basé sur l'utilisation de LexA, répresseur bactérien. LexA sauvage reconnaît spécifiquement et sous forme de dimères une séquence palindromique de 16 paires de bases, CTGT (N)₈ ACAG, appelée opérateur. Les 4 paires de bases externes sont fortement conservées, alors que le motif central est plus variable. Chaque monomère de LexA est composé d'un domaine de fixation à l'ADN qui reconnaît la séquence CTGT et d'un domaine de dimérisation indispensable à la fixation sur l'ADN puisque le

domaine de fixation à l'ADN dépourvu de domaine de dimérisation, ne fixe que très faiblement l'ADN.

La littérature décrit des mutants de LexA ayant une nouvelle spécificité de fixation à l'ADN. Celui qui a été utilisé, LexA408, porte une triple mutation (P40→A, N41→S, A42→S) et reconnaît la séquence suivante CCGT (N)₈ ACGG.

Le système selon la présente invention est basé sur le fait qu'un opérateur muté contenant 1/2 opérateur sauvage et 1/2 opérateur muté n'est reconnu que par un dimère mixte LexAWT/LexA408. Un gène indicateur (*lacZ*) également désigné par gène reporter a été mis sous le contrôle de cet opérateur. Des protéines ou des sous domaines de protéines X et Y susceptibles d'interagir sont ensuite fusionnés aux domaines de fixation à l'ADN, LexA₁₋₈₇WT et de LexA₁₋₈₇408. La figure 1 illustre ce système. LexA₁₋₈₇WT-X et LexA₁₋₈₇408-Y sont clonés sur des plasmides compatibles et leur expression est contrôlée par un promoteur inductible (*lacUV5*). La souche indicatrice surexprimant le répresseur du promoteur lactose (puisque *lacI^q*), on peut alors moduler l'expression de LexA₁₋₈₇WT-X et LexA₁₋₈₇408-Y en ajoutant des concentrations variables de l'inducteur IPTG.

Ce système a deux utilisations principales:

I)- X et Y peuvent être des protéines ou des domaines de protéines caractérisés et interagissants entre eux, ils confèrent alors aux protéines hybrides la capacité de se fixer de façon stable sur l'ADN et ainsi de réprimer le gène *lacZ*. Une telle interaction peut être visualisée sur un milieu indicateur approprié (par exemple Mac Conkey et décrit ci-après) et peut également être quantifiée par dosage du produit du gène *lacZ*, la β -galactosidase selon la méthode décrite par Miller J.H., 1972 "Experiments in Molecular Genetics, Cold Spring Harbor Lab. Cold Spring Harbor, NY) ; il est ainsi possible :

- 1) de quantifier la force de cette interaction ;
- 2) d'identifier les déterminants de cette interaction (par mutagenèse par exemple) ;
- 3) d'identifier des molécules capables d'interférer avec cette interaction et susceptibles d'être utilisées *in vivo* ; petites molécules susceptibles d'être mises dans le milieu de culture et de passer à travers la paroi bactérienne.

II)- X (ou Y) est une protéine ou un domaine de protéine susceptible d'interagir avec une ou des protéine(s) encore inconnue(s) que l'on veut identifier. On greffe alors sur Y (ou X) une banque, cette banque étant cointroduite dans la souche indicatrice avec le gène codant pour la protéine hybride dont on cherche le ou les partenaires. Tout élément de la banque codant

pour une protéine ou un domaine de protéine interagissant avec X (ou Y) donne lieu à un clone bactérien dont la couleur sur milieu approprié est différente de ceux pour lesquels il n'y a pas d'interaction. Ce ou ces éléments sont alors identifiés par séquençage.

EXEMPLE 1 :

Construction de la souche indicatrice pour la mise en oeuvre d'un système double hybride chez *E. coli*

Cette construction nécessite :

1) le clonage de *lacZ* sous le contrôle d'un promoteur régulé par un opérateur de type CCGT (N)g ACAG;

2) de faire passer cette construction sur le chromosome d'un hôte bactérien délété de son opéron lactose, déficient en *LexA* chromosomique et exprimant le répresseur lactose, souche JL1434 (Little J.W. & Hill S.A. 1985 *Deletions within a hinge region of a specific DNA-binding protein*. Proc. Natl. Acad. Sci. USA 82:2301-2305 et ceci par l'intermédiaire d'un phage lambda (Figure 2).

1. Construction de plasmides portant le gène *lacZ* placé sous le contrôle d'un opérateur muté WT / 408

Le plasmide pRS415 portant le gène *lacZ*, dépourvu de région promoteur/opérateur (Simons R.W., Houman F. & Kleckner N. (1987). *Improved single and multicopy lac-based cloning vectors for protein and operon fusions*. Gene 53:85-96.) est ouvert aux sites de restriction EcoRI/BamHI. Entre ces deux sites uniques de restriction, a été inséré l'oligonucléotide double brin suivant:

5'-GAATTCAATAGGGTTGATCTTTGTTGTCACTGGATGTACCGTACATCCATACAGTAACTCACAGGGGCACCCTAGG-3'
(bases en caractère gras = éléments du promoteur, bases soulignées = opérateur) créant ainsi le plasmide pGC202.

2. Isolement de bactériophages Lac⁺ recombinants

La souche JL1434 (*LexA*(Def), *Lac*⁻), transformée par les plasmides pGC202, a été infectée par le bactériophage $\lambda\phi$ d (*Lac*⁻, Hochschild A. & Ptashne M. (1988). *Interaction at a distance between λ repressors disrupts gene activation*. Nature 336:353-357) dont le génome porte immédiatement à gauche d'un gène *lacZ* dépourvu de promoteur et de codon d'initiation (ATG) de la traduction, une séquence dérivant du vecteur pBR322 et que l'on retrouve également à gauche du site de clonage EcoRI du plasmide pRS415 (voir figure

2). Grâce aux séquences homologues situées de part et d'autre de la séquence "promoteur/opérateur/initiateur" et suite à un double événement de recombinaison entre l'ADN plasmique et phagique, le gène *lacZ* se retrouve sous le contrôle de la région promoteur/opérateur cloné dans pGC202 donnant lieu au phage λ GC 202.

3. Isolement de bactéries lysogènes pour les bactériophages Lac⁺ recombinants

L'isolement de bactéries lysogènes a pour but d'obtenir l'intégration stable du phage recombinant dans le chromosome bactérien, qui s'y trouve alors à l'état de prophage. Ceci se fait à partir d'une deuxième infection des bactéries JL1434 avec des minipréparations du phage précédemment obtenu afin de se débarrasser du plasmide initial et de s'assurer ainsi que le caractère Lac⁺ est bien porté par le phage.

Après infection de la souche JL1434, les phages λ recombinants commencent un cycle lysogène qui conduit à l'intégration stable par recombinaison site-spécifique de leurs génomes (et donc du gène *lacZ* contrôlés par l'élément promoteurs/opérateurs décrit ci-dessus) dans le chromosome bactérien, au niveau du site *att b* entre les opérons *gal* et *bio*. La souche lysogène Lac⁺ SU202 a ainsi été isolée. Cette souche indicatrice porte les gènes de résistance à la Kanamycine et au Chloramphénicol.

Description des plasmides porteurs des protéines hybrides

Les gènes codant pour les protéines hybrides sont portés par deux vecteurs plasmidiques compatibles portant des origines de réplication différentes (pMB1 ou p15A) ainsi que la sélection pour deux marqueurs différents (tet^R ou amp^R) rendant ainsi possible leur maintien simultané dans la même bactérie hôte.

Utilisation de la souche SU202 pour identifier et quantifier des interactions protéine/protéine

La souche indicatrice SU202 a été testée avec les protéines chimères suivantes : LexA1-87408-cJunZip et LexA1-87WT-cFosZip, elles sont constituées du domaine de fixation à l'ADN (mutant 408 ou sauvage (WT)) du répresseur LexA (aa 1-87) et du domaine de dimérisation (glissière à leucine ou "leucine zipper") des oncoprotéines cJun (aa 277-315) et cFos (aa 102-200).

Le gène codant pour LexA1-87408-cJunZip est porté par le plasmide pDL606 (voir figures 6 et 7). La figure 6 représente une carte de l'ensemble du plasmide accompagné des sites uniques de restriction, le vecteur est un

pACYC177 (*New England Biolabs, Inc.*) ; il porte la résistance à l'ampicilline et p15A comme origine de réplication. La figure 7 représente la partie de la séquence de ce plasmide correspondant à la protéine de fusion, cette séquence étant repérable sur la carte générale à l'aide des sites de restriction.

Le gène codant pour LexA₁-87WT-cFosZip est porté par le plasmide pMS404 (voir figures 8 et 9). La figure 8 représente une carte de l'ensemble du plasmide accompagné des sites uniques de restriction, le vecteur est un plasmide apparenté à un pBR322 (*New England Biolabs, Inc.*) ; il porte la résistance à la tétracycline et pMB1 comme origine de réplication. La figure 9 représente la partie de la séquence de ce plasmide correspondant à la protéine de fusion, cette séquence étant repérable sur la carte générale à l'aide des sites de restriction.

Ces domaines de dimérisation sont tels que les homodimères Jun/Jun sont beaucoup moins stables que les hétérodimères Jun/Fos, mais cependant davantage que les homodimères Fos/Fos qui n'existent pas *in vivo* (Sassone-Corsi P., Ransone L.J., Lamph W.W. & Verma I.M. (1988). *Direct interaction between fos and jun nuclear oncoproteins: role of the "leucine-zipper" domain*. Nature (London) 334:692-695.).

Les gènes codant pour ces deux protéines de fusion sont contrôlés par le promoteur inductible *lacUV5*. Les souches indicatrices surexprimant ce répresseur (puisque *lacI^q*), permettent de moduler l'expression de LexA₁-87408-cJunZip et LexA₁-87WT-cFosZip en présence de différentes concentrations en inducteur (IPTG).

Test qualitatif

La souche SU 202 est transformée soit avec chacun des plasmides portant les gènes codant pour les protéines de fusion LexA₁-87408-cJunZip et LexA₁-87WT-cFosZip soit cotransformée avec les deux plasmides. Les cellules transformées sont laissées une nuit à température ambiante dans du LB (Maniatis *et al.*) additionné respectivement soit de kanamycine, chloramphénicol, ampicilline soit de kanamycine, chloramphénicol, tétracycline soit des quatre antibiotiques: kanamycine, chloramphénicol, ampicilline et tétracycline, avant d'être étalées sur milieu McConkey contenant les mêmes antibiotiques. La figure 5 montre, après repiquage sur une même boîte, les résultats obtenus.

Ce type de milieu de culture contient entre autre du lactose qui est hydrolysé par la β -galactosidase en allolactose. L'allolactose étant l'inducteur naturel du promoteur lactose, lorsque les bactéries possédant l'un ou les deux gènes codant pour les protéines chimères sont cultivées sur ce milieu, ceux-ci sont partiellement induits. On estime que l'induction correspond à celle que l'on

observe en milieu LB additionné de 10^{-5} M en IPTG. Il a été observé une différence importante dans la coloration des clones bactériens indiquant que la souche indicatrice construite répondait aux critères définis, en effet :

- les cellules transformées par un seul des deux plasmides poussent rouges; cette coloration signifie que le gène de la β -galactosidase n'est pas réprimé par les protéines chimères et que l'enzyme produit fermente le lactose du milieu; dans le cas LexA1-87408-cJunZip, les homodimères qui sont susceptibles de se former sont incapables *in vivo* de reconnaître l'opérateur muté et donc de réprimer le gène *lacZ*. En ce qui concerne la glissière à leucine ("leucine zipper") de Fos, la réponse était attendue puisqu'il n'homodimérise que très mal;

- les bactéries cotransformées par les deux plasmides poussent blanches, la β -galactosidase n'est donc pas produite par les cellules; cette double transformation permet à la cellule de produire simultanément les deux protéines chimères LexA1-87408-cJunZip et LexA1-87WT-cFosZip; l'absence de production de la β -galactosidase indique que l'hétérodimère LexA1-87408-cJunZip/LexA1-87WT-cFosZip, est contrairement aux homodimères, capable de réprimer le gène *lacZ*.

Cette souche indicatrice est donc bien fonctionnelle, elle permet de mettre en évidence des interactions protéine/protéine sur la simple base de la couleur des clones, elle est très spécifique et permet de discriminer entre homo- et hétérodimérisation.

Test quantitatif

L'efficacité avec laquelle le gène indicateur est réprimée peut être mesurée quantitativement par dosage *in vitro* de l'activité β -galactosidase. Ce test est effectué dans les conditions décrites par Miller J.H. (1972) dans *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Les valeurs des dosages décrits représentent la moyenne d'au moins trois expériences indépendantes et sont réalisées en présence de concentrations variables en IPTG.

La figure 3 montre que:

- * quelle que soit la concentration en IPTG, la souche non transformée produit le même nombre d'unités β -galactosidase (en moyenne 2500 unités);
- * lorsqu'elle est transformée avec le plasmide codant pour LexA1-87WT-cFosZip et quelle que soit la concentration en IPTG (c'est-à-dire même en surexprimant la protéine chimère), le nombre d'unités ne varie pas;

* lorsqu'elle porte les deux protéines chimères, on observe une répression très forte du gène reporter (le nombre d'unités β -galactosidase chute à 20 pour les concentrations 10^{-4} et 10^{-3} M en IPTG);

* cette répression est due à l'interaction spécifique entre les parties glissière à leucine ("leucine zippers") des protéines chimères aboutissant à la formation d'hétérodimères actifs, et non pas à la simple coexistence de deux protéines chimères dont les domaines de liaison à l'ADN seraient capables de se fixer indépendamment sur chaque demi site de l'opérateur muté; en effet, lorsque la souche porte le gène codant pour le domaine de fixation à l'ADN WT de LexA dépourvu de son domaine de dimérisation et celui codant pour LexA₁₋₈₇₄₀₈-cJunZip, on n'observe pas de répression du gène reporter.

On dispose donc d'un test permettant de mettre en évidence des interactions protéine/protéine utilisable comme crible mais aussi avec la possibilité de mesurer et de comparer la force de ces interactions.

Recherche et identification de mutants affectés dans ces interactions

Cette souche peut être également utilisée pour identifier les déterminants de ces interactions soit par mutagenèse aléatoire soit par mutagenèse dirigée. Par exemple, il avait été émis l'hypothèse que la spécificité d'hétérodimérisation entre les glissières ("zippers") de Jun et de Fos était due à la non homodimérisation de la glissière ("zipper") de Fos. Cette souche a permis de montrer que tel n'était pas le cas puisque de mutants de Fos capables d'homodimériser n'ont pas perdu cette spécificité (figure 4).

Les cellules transformées uniquement par les plasmides codant pour les mutants de Fos poussent rouges sur milieu McConkey. L'homodimérisation n'est donc pas détectée dans cette souche.

Dans le cas où les bactéries produisent en plus LexA₁₋₈₇₄₀₈-cJunZip, elles poussent blanches si les mutants de Fos sont améliorés dans leur interaction protéine/protéine par rapport à Fos sauvage, et roses si les mutants de Fos sont moins bons. L'intensité de la couleur est toujours inversement proportionnelle à la capacité d'hétérodimérisation (dosage β -galactosidase).

Utilisation de la souche SU202 pour identifier le ou les partenaires d'une protéine ou d'un domaine protéique caractérisé

Deux plasmides sont utilisés:

1- le plasmide portant la protéine ou le domaine de protéine caractérisé pour lequel on veut identifier le ou les partenaires fusionné(s) au domaine de fixation à l'ADN de LexA₄₀₈, pDL606 : constituant "l'hameçon".

2- le plasmide portant la banque, pLE23B (voir figures 10 et 11). La figure 10 représente une carte de l'ensemble du plasmide accompagné des sites uniques de restriction, le vecteur est un plasmide apparenté à un pBR322 (*New England Biolabs, Inc.*) il porte la résistance à la tétracycline et pMB1 comme origine de réplication. La figure 11 représente la partie de la séquence de ce plasmide correspondant à la protéine de fusion, cette séquence étant repérable sur la carte générale à l'aide des sites de restriction.

Les deux plasmides sont simultanément introduits dans la souche SU202 et les bactéries transformées sont :

- soit directement sélectionnées sur milieu solide LB + antibiotiques (kanamycine, chloramphénicol, ampicilline et tétracycline), les clones positifs étant ensuite révélés par répliques des boîtes sur milieu indicateur McConkey + les mêmes antibiotiques,

- soit sélectionnées en milieu liquide LB + antibiotiques (kanamycine, chloramphénicol, ampicilline et tétracycline) pendant une nuit, puis révélées par étalement des cultures sur milieu indicateur McConkey + les mêmes antibiotiques.

Chaque clone porte un élément de banque différent codant pour une protéine ou un domaine de protéine différent, lorsque cette protéine ou ce domaine de protéine est susceptible d'établir des interactions protéine/protéine avec "l'hameçon", il donne lieu à un clone blanc ou rosé parmi un tapis de clones rouges, qui eux représentent tous les éléments de la banque codant pour des protéines ou des domaines de protéine n'interagissant pas avec l'hameçon.

La constitution de la banque passe par l'intermédiaire d'une navette phage Elledge S.J. *et al.*, 1991, Proc. Natl. Acad. Sci. USA, 88:1731-1735. " λ Yes: multifunctional cDNA expression vector for the isolation of genes by complementation of yeast and *E. coli* mutations".

EXEMPLE 2 :

Construction du système de sélection mettant en oeuvre un gène indicateur codant pour une protéine toxique :

Modification de la séquence opérateur du gène *sulA* de manière à rendre son expression dépendante d'interactions hétérodimériques entre des protéines ou domaines protéiques X et Y fusionnés respectivement aux domaines de fixation à l'ADN de LexA408 et LexAWT.

Cette modification consiste à introduire une mutation ponctuelle dans la séquence opérateur naturelle de *sulA* :

CTGTACATCCATACAG = opérateur WT,

CCGTACATCCATACAG = opérateur mixte 408/WT.

Le gène *sulA* est *in vivo* réprimé par LexA, l'introduction de la mutation op408 empêche cette répression, le produit du gène est alors synthétisé et les bactéries meurent, rendant impossible la sélection d'une telle construction. Il est donc nécessaire de trouver un moyen autre que la répression pour contrôler la production de SULA.

Préalablement à l'instauration de cette mutation, on a introduit de deux à quatre codons STOP dans la phase codante de *sulA* (un codon STOP est une séquence spécifique de trois paires de bases servant de signal d'arrêt à la machinerie traductionnelle). Ainsi, seule une protéine tronquée inactive est produite et ceci permet l'instauration de la mutation op408.

Lorsque le système est mis en oeuvre, il suffit d'utiliser un hôte bactérien porteur d'un ARNt suppresseur de ce codon STOP (un ARNt suppresseur est un ARNt capable de reconnaître un codon STOP et de permettre la poursuite de la traduction). L'hôte bactérien (SU202) que l'on utilise étant porteur de l'ARNt suppresseur SupE qui, à la place des codons STOP (UAG), introduit un résidu Glutamine, on a remplacé de deux à quatre résidus glutamine par des codons TAG.

Le gène *sulA* cloné dans un plasmide a été donné par S. Cole (Institut Pasteur, Paris). On a introduit les codons STOP TAG au niveau des Glutamines 42,47 d'une part, et 42,47 55 et 56 d'autre part, donnant lieu à des constructions appelées respectivement *sulA 2STOP* et *sulA 4STOP*. Ceci a été réalisé par mutagenèse dirigée selon la méthode décrite par Eckstein (Nakamaye, K. et Eckstein, F. Nucleic Acids Res. 1986, 14:9679-9698) et en utilisant un kit commercialisé par la société Amersham.

La mutation 408 a ensuite été introduite par le même procédé de mutagenèse et des résultats identiques ont été obtenus avec *sulA 2STOP* et *sulA 4STOP*. Les constructions obtenues ont été appelées *op408/opWT sulA 2STOP* et *op408/opWT sulA 4STOP*.

Mise en oeuvre du système de sélection ;

Le système a été testé dans la souche SU202 portant un ARNt suppresseur, SupE. Les résultats suivants ont été observés.

i) *op408/opWT sulA 2STOP* ou *op408/opWT sulA 4STOP* ne peuvent pas être introduits dans la souche. En effet, quelle que soit la méthode de transformation utilisée (chlorure de calcium ou électroporation), aucun clone n'a été obtenu, alors qu'un plasmide identique mais ne portant pas la construction

donne lieu à 200 ou 2000 clones par nanogramme de plasmide selon la méthode de transformation utilisée ;

ii) par contre, *op408/opWT sula 2STOP* ou *op408/opWT sula 4STOP* peuvent être introduits dans la souche SU202 s'ils sont co-transformés avec des plasmides portant les gènes codant pour les protéine de fusion LexA₁₋₈₇408-cJunZip et LexA₁₋₈₇WT-cFosZip.

Dans la première expérience, l'opération de *sula* n'est pas occupé et les codons STOP sont "lus" par l'ARNt suppresseur, SupE. La protéine SULA est produite et provoque la mort de la bactérie.

Dans la deuxième expérience, les hétérodimères LexA₁₋₈₇408cJunZip / LexA₁₋₈₇WT-cFosZip sont capables de se fixer sur l'opérateur de *sula* et empêchent ainsi la production de la protéine SULA, la bactérie peut se diviser et donner lieu à des clones.

REFERENCES

1. Bartel, P.L., Chien, C.-T., Sternglanz, R. & Fields, S. (1993) "Using the two-hybrid system to detect protein-protein interactions", In: *Cellular Interactions in Development: A practical approach* (Oxford University Press, Oxford) pp. 153-179.
2. Chien, C.-T., Bartel, P.L., Sternglanz, R. & Fields, S. (1991), "The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest", *Proc. Natl. Acad. Sci. USA* **88**:9578-9582.
3. Durfee, T., Becherer, K., Chen, P.-L., Yeh, S.-H., Yang, Y., Kilburn, A.E., Lee, W.-H. & Elledge, S.J. (1993), "The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit", *Genes and Dev.* **7**:555-569.
4. Estojak, J., Brent, R. & Golemis, E.A. (1995), "Correlation of two-hybrid affinity data with in vitro measurements", *Mol. Cell. Biol.* **15**:5820-5829.
5. Fields, S. & Song, O.-K. (1989) "A novel genetic system to detect protein-protein interactions", *Nature* **340**:245-246.
6. Fogh, R.H., Otteleben, G., Rüterjans, H., Schnarr, M., Boelens, R. & Kaptein, R. (1994), "Solution structure of the LexA repressor DNA binding domain: determination by 1H NMR spectroscopy", *EMBO J.* **13**:3936-3944.
7. Golemis, E.A. & Brent, R. (1992), "Fused protein domains inhibit DNA binding by LexA", *Mol. Cell. Biol.* **12**:3006-3014.
8. Golemis, E.A. & Brent, R. (1996), "The interaction trap", In: *Two-hybrid systems: A practical approach* (Oxford University Press, Oxford) Chapter 4.
9. Golemis, E.A., Gyuris, J. & Brent, R. (1996), "Interaction trap/two-hybrid system to identify interacting proteins", In: *Current Protocols in Molecular Biology*, John Wiley & Sons, New York (Eds: Ausubel, F., Brent,

- R., Kingston, R., Moore, D., Seidmann, J., Smith, J.A. & Struhl, K.) Unit 20.1.
10. Gyuris, J., Golemis, E.A., Chertkov, H. & Brent, R. (1993) "Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2", *Cell* **75**:791-803.
 11. Hori, R. & Carey, M. (1994) "The role of activators in assembly of RNA polymerase II transcription complexes", *Current Opinion in Genetics and Dev.* **4**:236-244.
 12. Hu, J.C. (1995) "Repressor fusions as a tool to study protein-protein interactions", *Structure* **3**:431-433.
 13. Hurstel, S., Granger-Schnarr, M., Daune, M. & Schnarr, M. (1986), "In vitro binding of LexA repressor to DNA: evidence for the involvement of the amino-terminal domain", *EMBO J.* **5**:793-798.
 14. Le Douarin, B., Pierrat, B., vom Baur, E., Chambon, P. & Losson, R. (1995) "A new version of the two-hybrid assay for detection of protein-protein interactions", *Nucl. Acids Res.* **23**:876-878.
 15. Llobès, R., Granger-Schnarr, M., Lazdunski, C. & Schnarr, M. (1991), "Interaction of a regulatory protein with a DNA target containing two overlapping binding sites", *J. Biol. Chem.* **266**:2303-2312.
 16. Ma, J. & Ptashne, M. (1987a), "Deletion analysis of Gal4 defines two transcriptional activating segments", *Cell* **48**:847-853.
 17. Ma, J. & Ptashne, M. (1987b), "A new class of yeast transcriptional activators", *Cell* **51**:113-119.
 18. Ma, J. & Ptashne, M. (1988), "Converting a eucaryotic transcriptional inhibitor into an activator", *Cell* **55**:443-446.
 19. Maldonado, E. & Reinberg, D. (1995), "News on initiation and elongation of transcription by RNA polymerase II", *Current Opinion in Cell Biol.* **7**:352-361.

20. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S.C. (1992), "DNA recognition by Gal4: structure of a protein-DNA complex", *Nature* 356:408-414.
21. Oertel-Buchheit, P., Lamerichs, R., Schnarr, M. & Granger-Schnarr, M. (1990), "Genetic analysis of the LexA repressor: isolation and characterization of LexA (Def) mutant proteins", *Mol. Gen. Genet.* 223:40-48.
22. Phizicky, E.M. & Fields, S. (1995), "Protein-protein interactions: methods for detection and analysis", *Microbiol. Rev.* 59:94-123.
23. Porte, D., Oertel-Buchheit, P., Granger-Schnarr, M. & Schnarr, M. (1995), "Fos leucine zipper variants with increased association capacity", *J. Biol. Chem.* 270:22721-22730.
24. Schmidt-Dörr, T., Oertel-Buchheit, P., Pernelle, C., Bracco, L., Schnarr, M. & Granger-Schnarr, M. (1991), "Construction, purification and characterization of a hybrid protein comprising the DNA binding domain of the LexA repressor and the Jun leucine zipper", *Biochemistry* 30:9657-9664.
25. Schnarr, M. & Granger-Schnarr, M. (1993), "LexA, the self-cleaving transcriptional repressor of the SOS system", *Nucleic Acids and Mol. Biol.* 7:170-189 (Eds. F. Eckstein & D.M.J. Lilley)
26. Schnarr, M., Pouyet, J., Granger-Schnarr, M. & Daune, M. (1985), "Large-scale purification, oligomerization equilibria, and specific interaction of the LexA repressor of *Escherichia coli*", *Biochemistry* 24:2812-2818.
27. Vojtek, A.B., Hollenberg, S.M. & Cooper, J. A. (1993), "Mammalian Ras interacts directly with the serine/threonine kinase Raf", *Cell* 74:205-214.

REVENDICATIONS

1- Utilisation d'une séquence d'ADN comprenant :

- une séquence nucléotidique codant pour une protéine indicatrice (gène indicateur),
- un promoteur,
- un opérateur muté,

ladite séquence nucléotidique étant sous le contrôle du susdit promoteur comprenant le susdit opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte une mutation 2, les homodimères susceptibles d'être formés à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

- pour détecter ou quantifier des interactions entre deux protéines ou domaines protéiques, ou cribler des protéines ou domaines protéiques présentant des interactions avec une protéine ou domaine protéique déterminé(e), ces interactions étant révélées lorsqu'il y a répression de la transcription du susdit gène indicateur, cette répression résultant de la formation d'hétérodimères comprenant :

1) un premier monomère contenant une première protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence de l'opérateur muté contenant la mutation 1, le susdit site de fixation étant fusionné à l'une des deux protéines ou à l'un des deux domaines protéiques, impliqué(e) dans les susdites interactions,

2) un deuxième monomère contenant une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon

silencieuse ou comportant la mutation 2 de l'opérateur muté, le susdit site de fixation étant fusionné à l'autre des deux protéines ou à l'autre des deux domaines protéiques, impliqué(e) dans les susdites interactions.

2. Séquence d'ADN comprenant :

- une séquence nucléotidique codant pour une protéine indicatrice (gène indicateur),

- un promoteur,

- un opérateur muté,

ladite séquence nucléotidique étant sous le contrôle d'un promoteur comprenant un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte une mutation 2, les homodimères susceptibles d'être formés à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

ladite mutation 1 étant en outre telle qu'elle est reconnue spécifiquement par une protéine régulatrice comportant un site de fixation à l'ADN entrant dans la constitution d'une première protéine de fusion, lequel site de fixation est fusionné avec une protéine ou un domaine protéique susceptible d'interagir avec une autre protéine ou un autre domaine protéique entrant dans la constitution d'une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon silencieuse ou comportant la mutation 2.

3. Séquence d'ADN selon la revendication 2, caractérisée en ce qu'elle comprend un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), l'une de ces deux séquences étant mutée, et l'autre étant à l'état sauvage.

4. Séquence d'ADN selon l'une des revendications 2 ou 3, caractérisée en ce que l'opérateur muté contient les séquences suivantes :

CCGT et ACAG

espacées par une séquence d'environ 3 à environ 20 nucléotides, et notamment de 8 à 12, et avantageusement 8 nucléotides, et notamment caractérisée en ce que l'opérateur muté contient la séquence suivante :

5'-CGGAATTCAATAGGGTTGATCTTTGTTGTCACTGGATGT
ACCGTACATCCATACAGTAACTCACAGGGGC-3'

5. Séquence d'ADN selon l'une des revendications 3 à 4, caractérisée en ce que

la séquence nucléotidique codant pour la protéine indicatrice est le gène *lacZ* chez *E. coli*, codant pour la β -galactosidase,

l'opérateur muté contient deux séquences et dérive d'un opérateur contenant deux séquences répétées inversées (palindrome), l'une de ces deux séquences étant mutée, et l'autre étant à l'état sauvage, la séquence mutée étant de préférence celle située en aval de la séquence à l'état sauvage,

la séquence sauvage étant reconnue spécifiquement par la protéine LexA sauvage ou par un fragment de LexA contenant le site de fixation à l'ADN notamment un fragment défini par les positions 1 à x de LexA, x variant de 69 à 87, et notamment par les fragments 1 à 69 ou 1 à 87 de LexA,

et la séquence mutée étant reconnue par la protéine LexA mutée ou par un fragment muté de LexA contenant le site de fixation à l'ADN, notamment un fragment muté de LexA correspondant à celui défini par les positions 1 à x de LexA, x variant de 69 à 87, et notamment par les fragments 1 à 69 ou 1 à 87, et les mutations de LexA étant avantageusement P40 \rightarrow A, N41 \rightarrow S, A42 \rightarrow S, le domaine de dimérisation de LexA étant absent à la fois de LexA ou du fragment de LexA et de LexA muté ou du fragment muté de LexA.

6. Complexe entre

- une séquence d'ADN comprenant une séquence nucléotidique codant pour une protéine indicatrice, ladite séquence nucléotidique étant sous le contrôle d'un promoteur comprenant un opérateur muté contenant deux séquences et dérivant d'un opérateur contenant deux séquences répétées inversées (palindrome), cet

opérateur étant tel que, à l'état non muté (sauvage), il entraîne la répression de la transcription du gène indicateur, lorsque des protéines régulatrices sous forme d'homodimère se fixent sur chacune desdites séquences répétées inversées via leur site de fixation à l'ADN, cet opérateur muté étant tel que l'une de ses séquences est mutée (mutation 1), tandis que l'autre est à l'état sauvage, ou mutée de façon silencieuse ou comporte une mutation différente (mutation 2) de la susdite mutation 1, la transcription du gène indicateur étant réprimée par des protéines régulatrices sous forme d'hétérodimères via leur site de fixation à l'ADN, chaque monomère de ces hétérodimères étant tel qu'il reconnaisse spécifiquement l'un la séquence comportant la mutation 1, l'autre la séquence qui est à l'état sauvage, ou qui est muté de façon silencieuse ou qui comporte une mutation 2, les homodimères susceptibles d'être fournis à partir de chacun des monomères étant incapables de réprimer la transcription du susdit gène indicateur,

- et des hétérodimères dont les monomères les constituant (respectivement premier monomère et deuxième monomère) sont impliqués entre eux dans une interaction protéine/protéine :

1) le premier monomère étant tel qu'il contient une première protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence de l'opérateur muté contenant la mutation 1, le susdit site de fixation étant fusionné à l'une des deux protéines ou à l'un des deux domaines protéiques, impliqué(e) dans les susdites interactions,

2) le deuxième monomère étant tel qu'il contient une deuxième protéine de fusion contenant le site de fixation à l'ADN d'une protéine régulatrice susceptible de reconnaître spécifiquement la séquence non mutée, ou mutée de façon silencieuse ou comportant la mutation 2 de l'opérateur muté, le susdit site de fixation étant fusionné à l'autre des deux protéines ou à l'autre des deux domaines protéiques, impliqué(e) dans les susdites interactions.

7. Complexe selon la revendication 6, caractérisé en ce que :

- la séquence d'ADN est telle que définie selon l'une quelconque des revendications 3 à 5, et en ce que
- les hétérodimères contiennent :

1) un monomère contenant une première protéine de fusion contenant le site de fixation à l'ADN de LexA muté ou d'un fragment muté de LexA reconnaissant la séquence de l'opérateur muté contenant la mutation 1, lequel site de fixation est fusionné avec une protéine déterminée ou un fragment

protéique déterminé susceptible d'être impliqué dans une interaction selon la revendication 1, laquelle protéine ou lequel domaine protéique contient par exemple un domaine d'activation ou de dimérisation,

2) un monomère contenant une deuxième protéine de fusion contenant le site de fixation à l'ADN de LexA sauvage, ou porteur d'une mutation silencieuse ou d'une mutation 2 ou d'un fragment de LexA reconnaissant la séquence non mutée de LexA ou mutée de façon silencieuse ou comportant une mutation 2, lequel site de fixation est fusionné avec une protéine déterminée ou un fragment protéique déterminé, contenant par exemple une protéine ou un domaine protéique susceptible d'interagir avec la protéine ou le fragment protéique défini en 1) ou avec les susdits domaines d'activation ou de dimérisation.

8. Hôte cellulaire contenant une séquence d'ADN selon la revendication 2.

9. Procédé pour détecter et quantifier des interactions entre deux protéines ou fragments de protéines, caractérisé

- en ce que l'on met en présence des séquences nucléotidiques codant respectivement pour le premier et le deuxième monomère contenant respectivement une première et une deuxième protéine de fusion telle que définie dans la revendication 1, avec une séquence d'ADN selon l'une quelconque des revendications 2 à 5, et

- en ce que l'on détecte l'éventuelle répression de la transcription du gène indicateur tel que défini à la revendication 1, notamment par transfection d'un hôte cellulaire selon la revendication 8, avec des vecteurs contenant, d'une part, les susdites séquences nucléotidiques codant respectivement pour le premier et le deuxième monomère contenant respectivement la première et la deuxième protéine de fusion et, d'autre part, les éléments nécessaires à l'expression des susdits monomères.

10. Procédé pour identifier une protéine interagissant avec une protéine déterminée, caractérisé

- en ce que l'on met en présence les éléments suivants :

- . une séquence nucléotidique codant pour un premier monomère contenant une première protéine de fusion entre, d'une part, une protéine régulatrice selon la revendication 1 et, d'autre part, la protéine déterminée,

. une banque génomique, contenant notamment des fragments d'ADN, et notamment d'ADNc, chacun étant susceptible de coder pour une protéine interagissant avec la protéine déterminée, chacun des différents éléments de la banque génomique étant respectivement fusionné à la séquence codant pour une protéine régulatrice selon la revendication 1,

. une séquence d'ADN selon l'une quelconque des revendications 2 à 5, et

- en ce que l'on détecte l'éventuelle répression de la transcription de la séquence nucléotidique codant pour une protéine indicatrice selon la revendication 1, notamment par transfection d'un hôte cellulaire selon la revendication 8, d'une part, avec un vecteur contenant la séquence nucléotidique codant pour le susdit premier monomère contenant la première protéine de fusion et, d'autre part, un ensemble de vecteurs contenant une séquence nucléotidique codant pour un deuxième monomère contenant une deuxième protéine de fusion, chaque séquence nucléotidique codant pour chaque deuxième monomère contenant la séquence nucléotidique codant pour la protéine régulatrice selon la revendication 1, et en aval de celle-ci l'un des éléments d'ADN de la susdite banque, tous les vecteurs contenant en outre les éléments nécessaires à l'expression des susdits monomères.



2/12

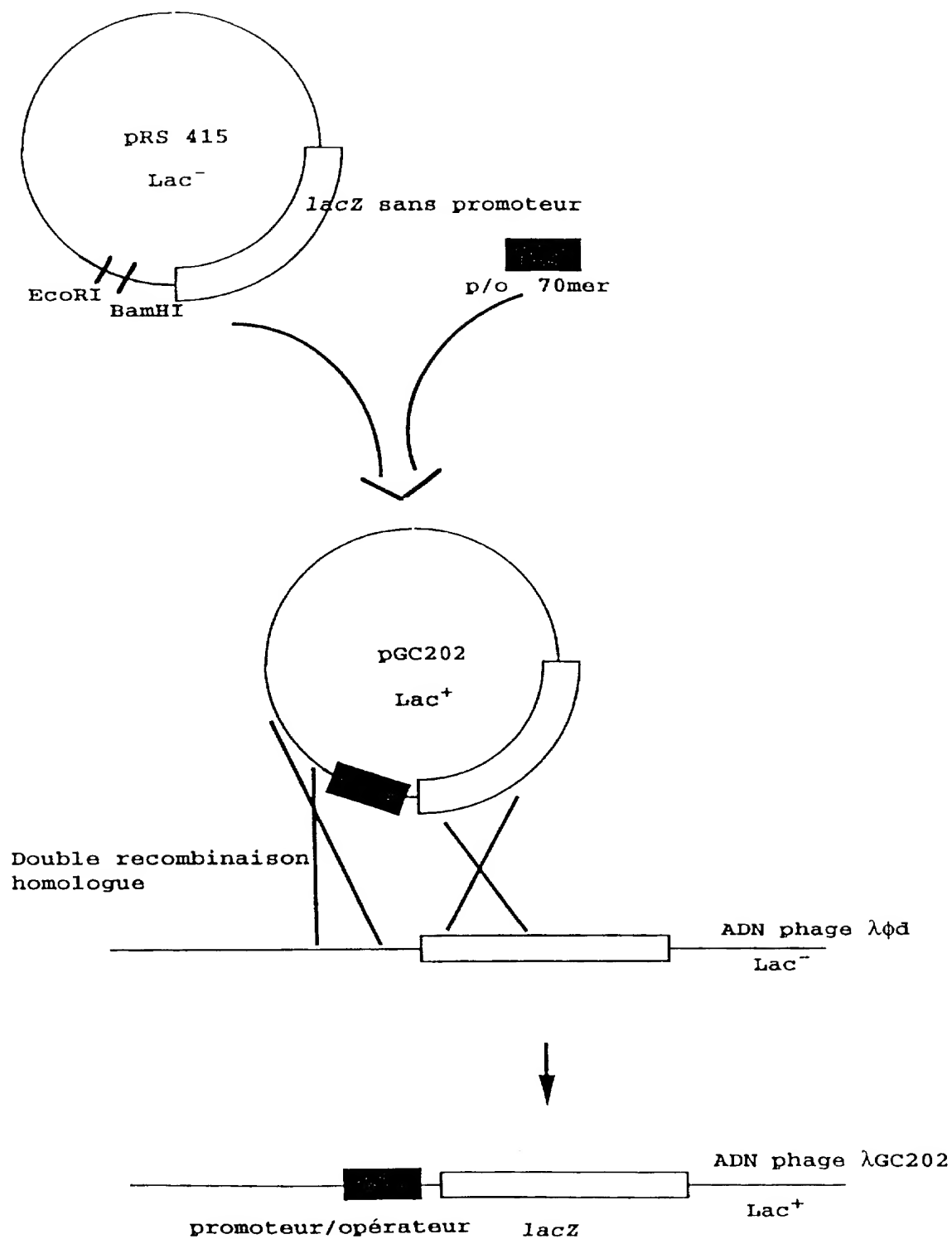


Figure 2

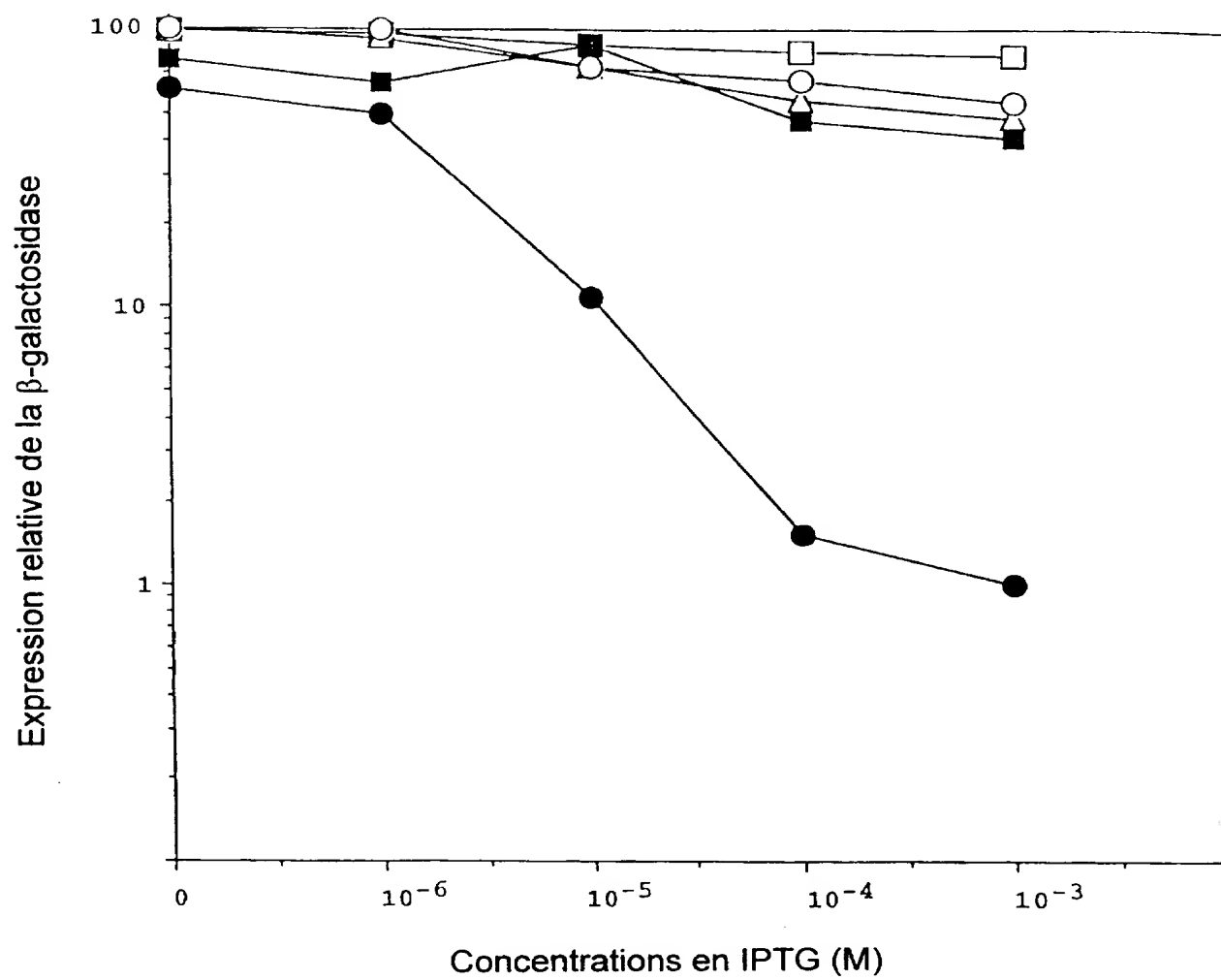


FIGURE 3

4/12

Figure 4 A

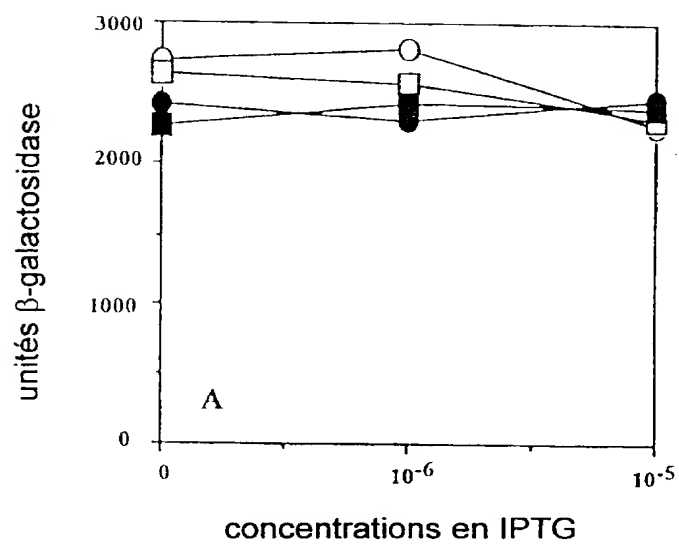


Figure 4B

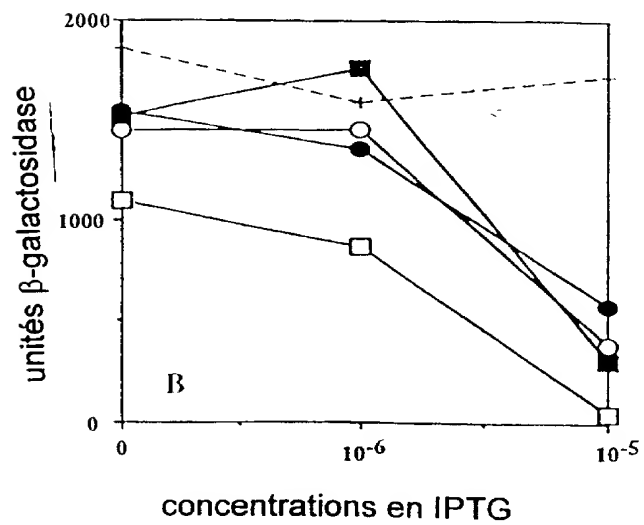
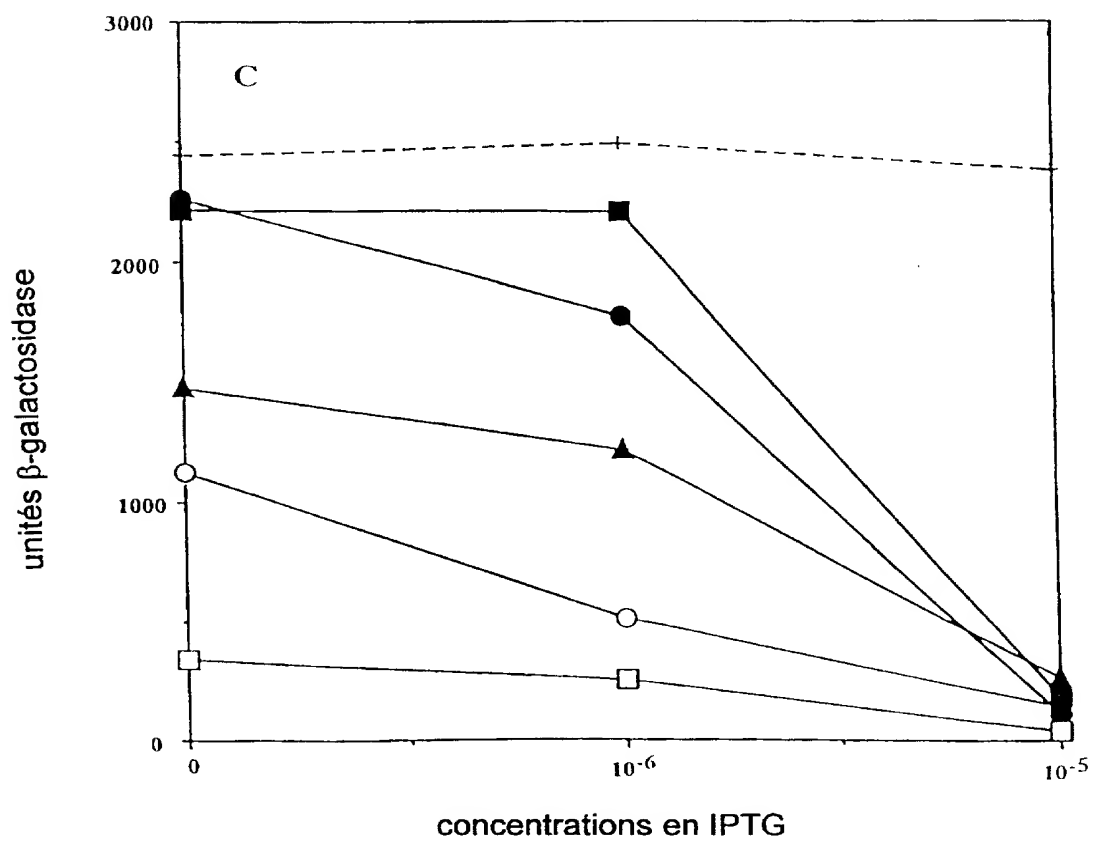


Figure 4 C



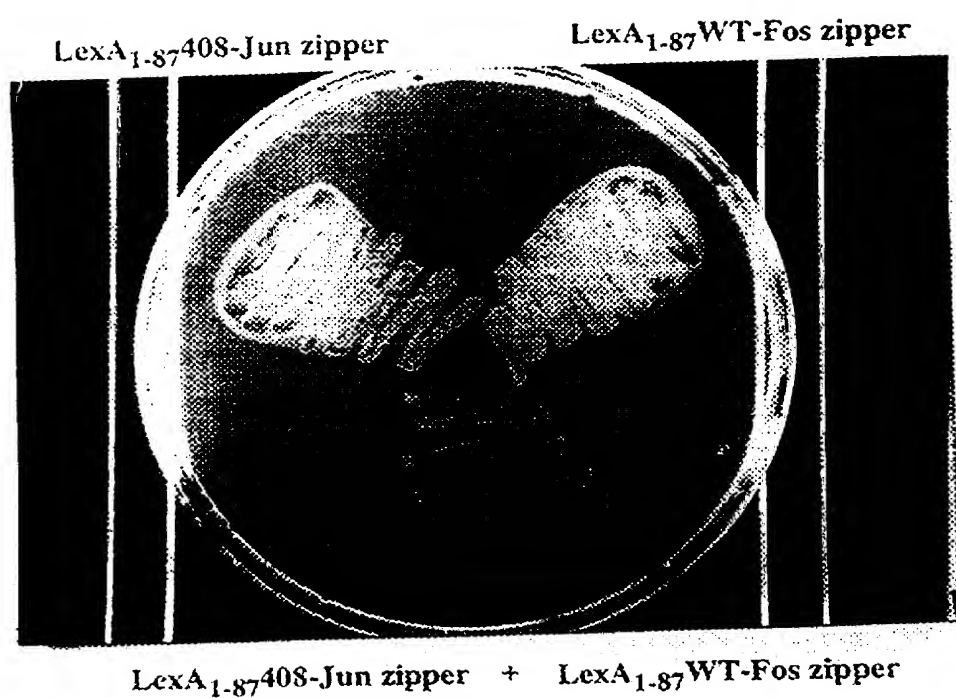


Figure 5

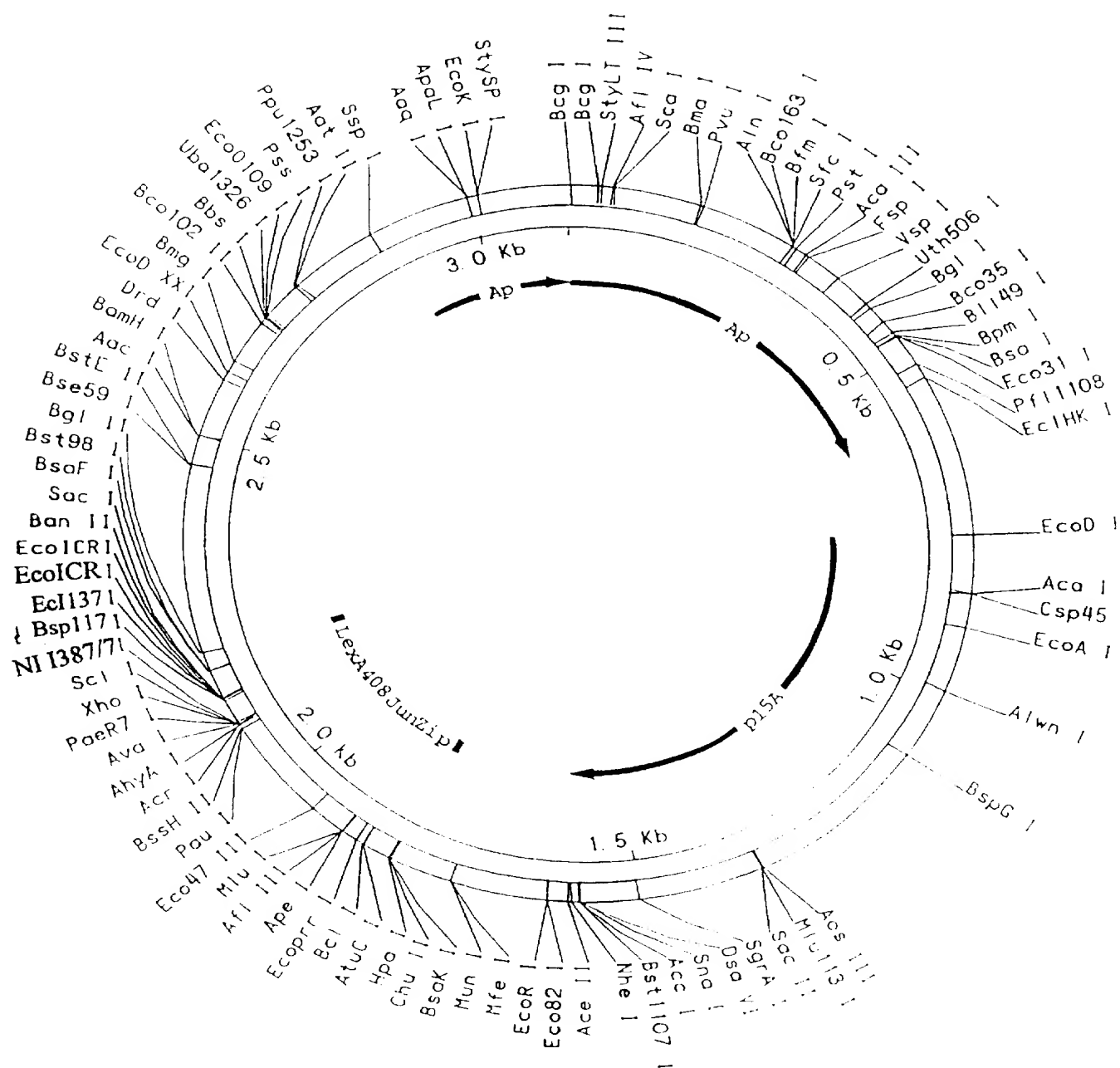


Figure 6

FIGURE 7

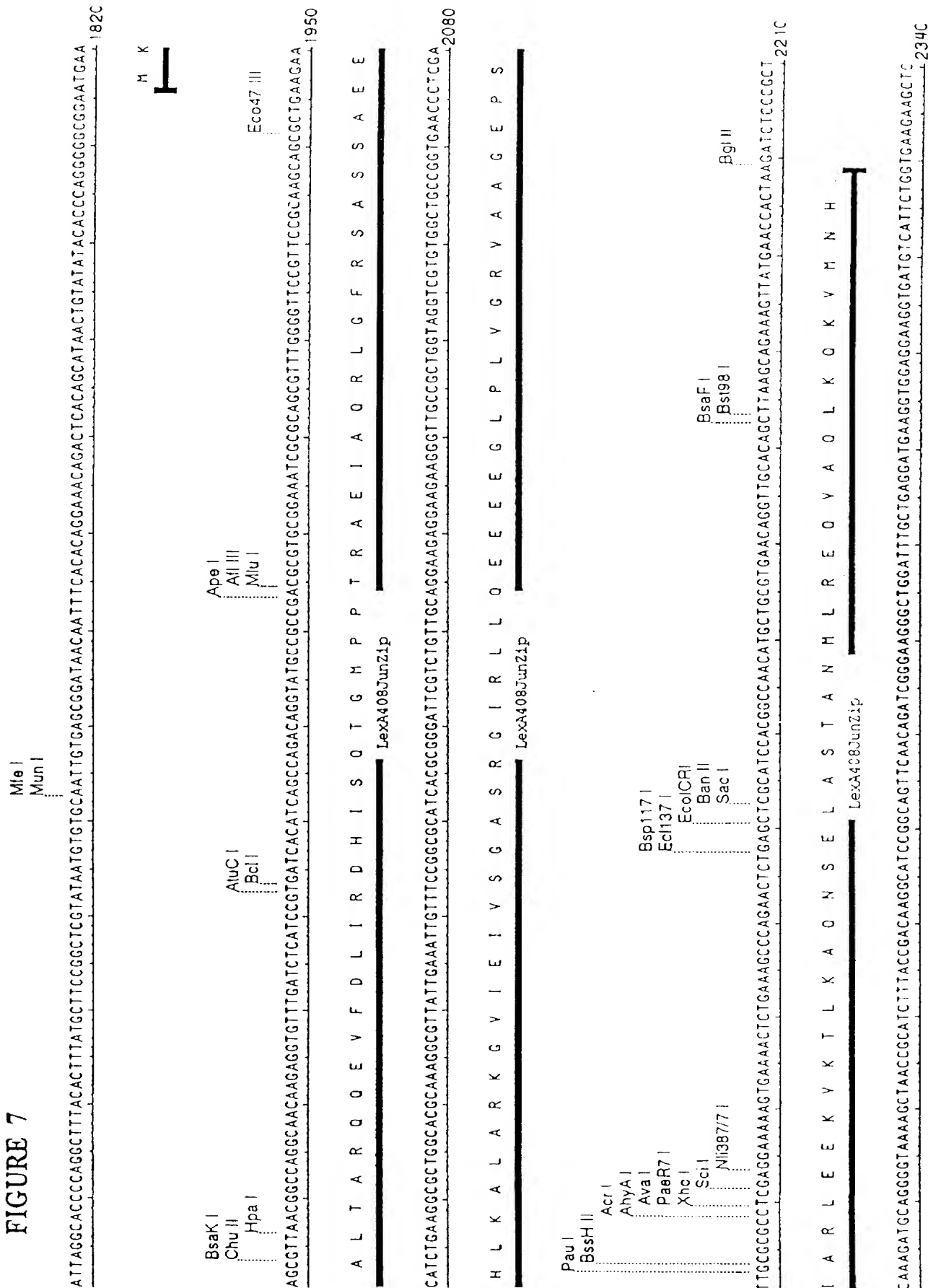
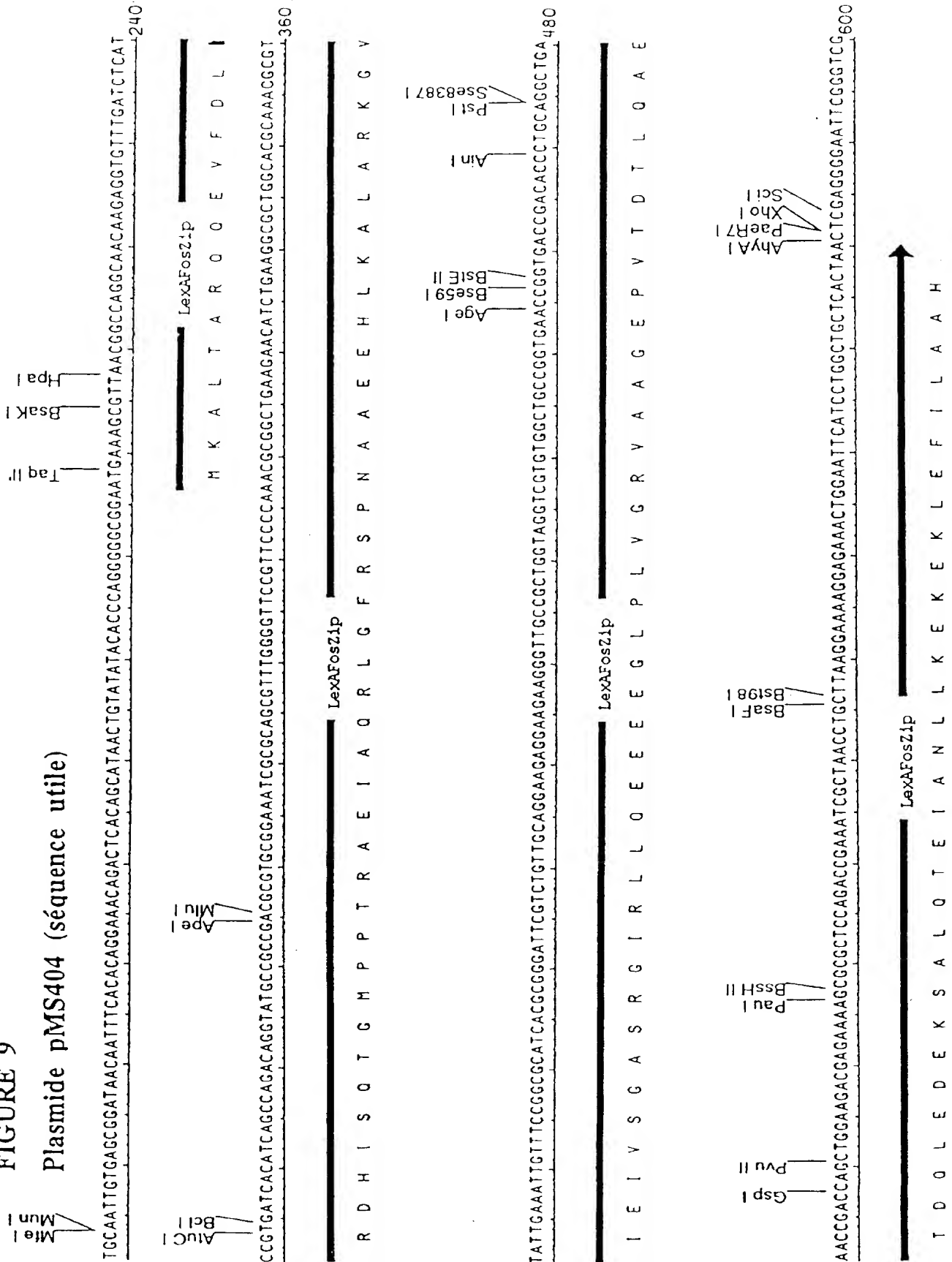


FIGURE 8

FIGURE 9

Plasmide pMS404 (séquence utile)



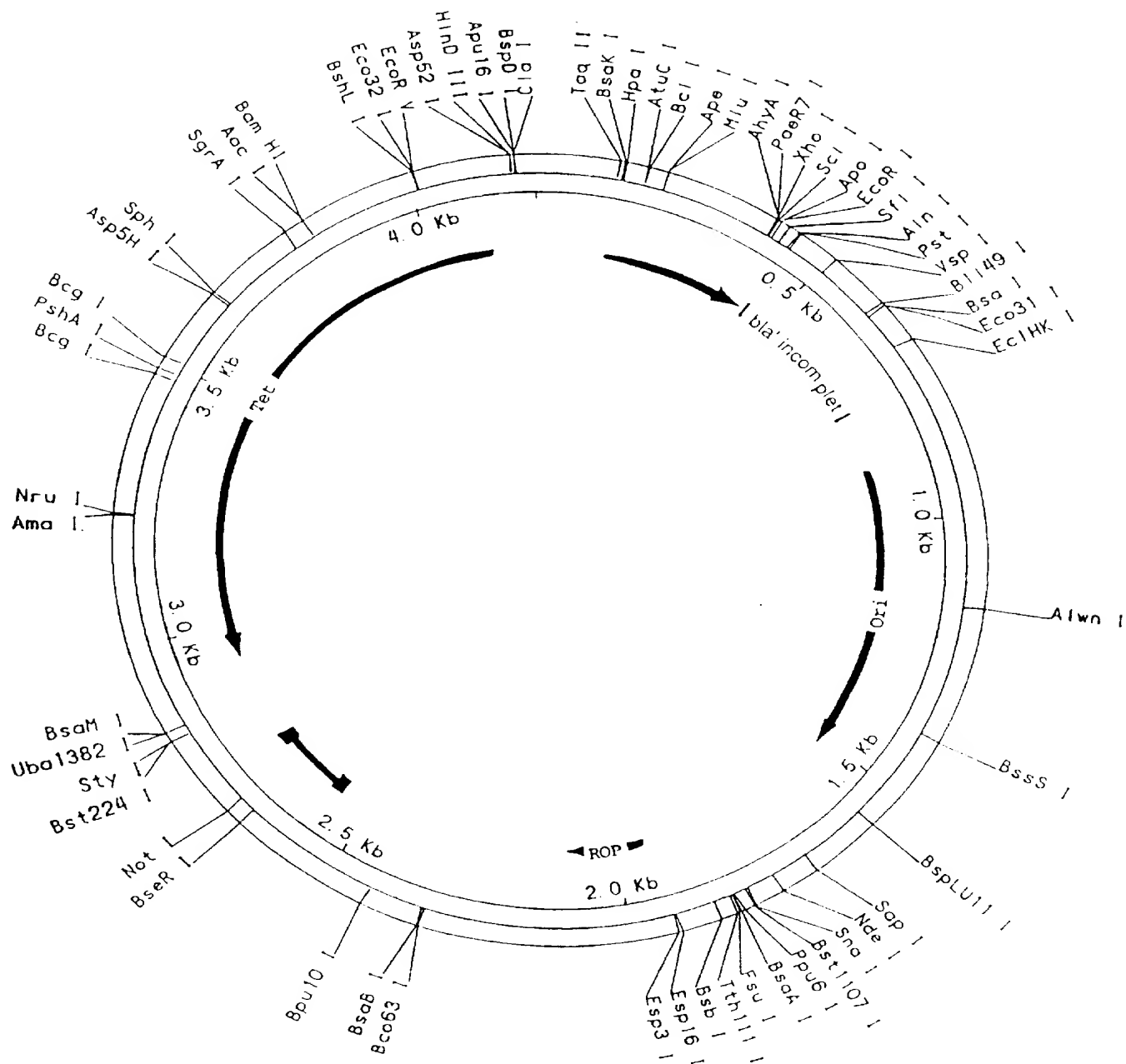
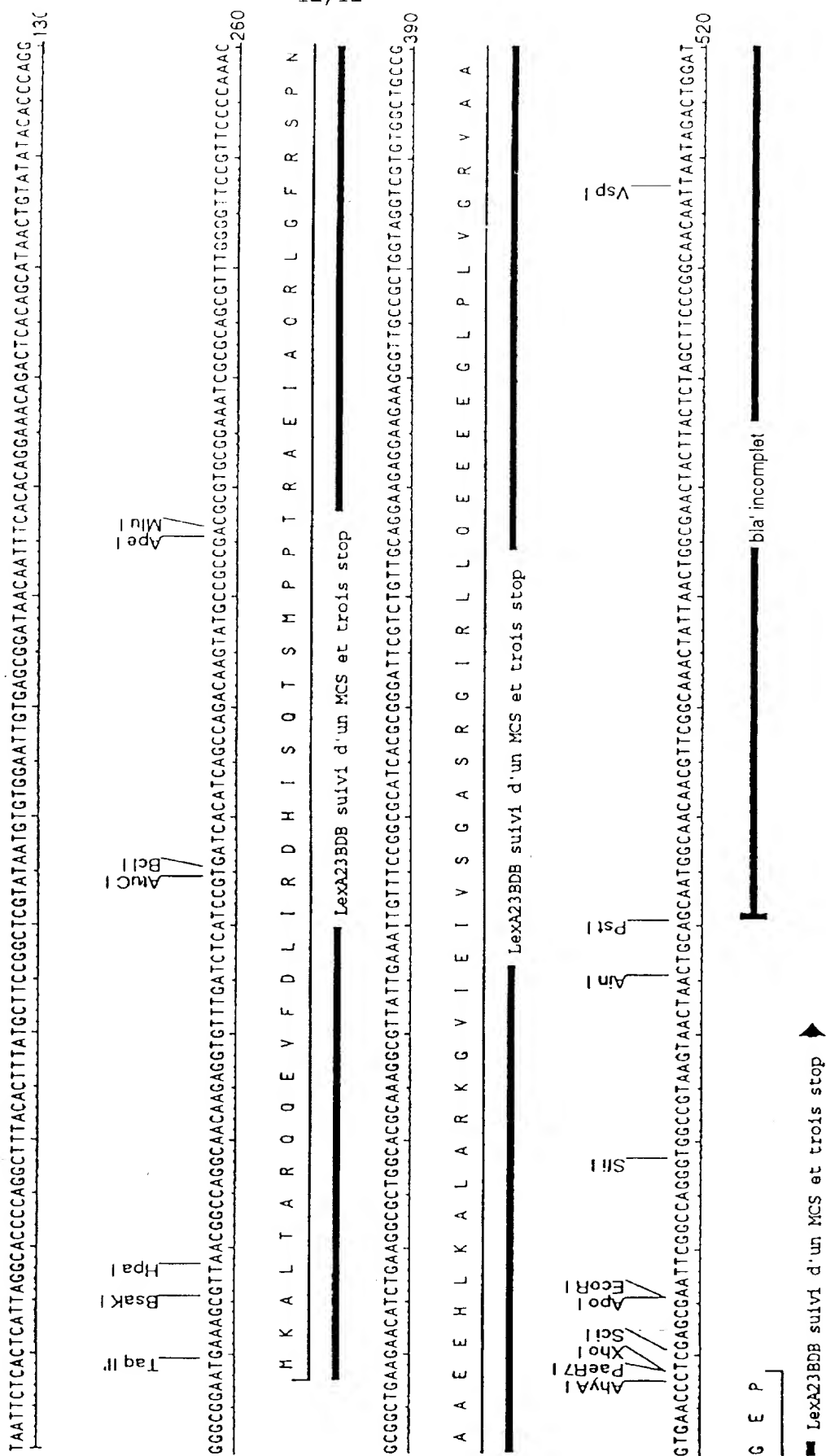


FIGURE 10

Plasmide pL23B (séquence utile)



INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 97/01133

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/68 C12N15/56 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 94 10307 A (MEDICAL RES COUNCIL ;THANGUE NICHOLAS BARRIE (GB)) 11 May 1994 see the whole document ---	1-10
A	US 5 283 173 A (FIELDS STANLEY ET AL) 1 February 1994 see the whole document ---	1,2,5,6, 8-10
A	COLE: "Characterisation of the promoter" MOLECULAR & GENERAL GENETICS, vol. 189, no. 3, 1983, BERLIN, pages 400-404, XP000197163 see figure 1 --- -/--	4

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

6 October 1997

Date of mailing of the international search report

22. 10. 97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Ceder, 0

INTERNATIONAL SEARCH REPORT

Intern: il Application No
PCT/FR 97/01133

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FEILDS ET AL.: "The two-hybrid system : an assay" TRENDS IN GENETICS, vol. 10, no. 8, August 1994, pages 286-292, XP000647708 cited in the application see the whole document ---	1-10
A	LITTLE ET AL.: "Deletions within a hinge region" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE USA, vol. 82, April 1985, WASHINGTON DC, pages 2301-2305, XP000647703 cited in the application see abstract ---	1
A	OERTEL-BUCHHEIT ET AL.: "Genetic analysis of the LexA repressor:" MOLECULAR & GENERAL GENETICS, vol. 223, 1990, BERLIN, pages 40-48, XP000647715 cited in the application see abstract -----	1

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demar: ternationale No

PCT/FR 97/01133

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9410307 A	11-05-94	AU 5343994 A	24-05-94
		EP 0669976 A	06-09-95
		JP 8503128 T	09-04-96
		NO 951641 A	29-06-95
		NZ 257181 A	27-07-97

US 5283173 A	01-02-94	US 5468614 A	21-11-95
		US 5667973 A	16-09-97

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No
PCT/FR 97/01133

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 C12Q1/68 C12N15/56 C12N1/21

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 6 C12Q

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	WO 94 10307 A (MEDICAL RES COUNCIL ;THANGUE NICHOLAS BARRIE (GB)) 11 mai 1994 voir le document en entier ---	1-10
A	US 5 283 173 A (FIELDS STANLEY ET AL) 1 février 1994 voir le document en entier ---	1,2,5,6, 8-10
A	COLE: "Characterisation of the promoter" MOLECULAR & GENERAL GENETICS, vol. 189, no. 3, 1983, BERLIN, pages 400-404, XP000197163 voir figure 1 --- -/-	4

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- "E" document antérieur, mais publié à la date de dépôt international ou après cette date
- "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"Z" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

6 octobre 1997

Date d'expédition du présent rapport de recherche internationale

22. 10. 97

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Ceder, 0

RAPPORT DE RECHERCHE INTERNATIONALE

Deman internationale No
PCT/FR 97/01133

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Calégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	FEILDS ET AL.: "The two-hybrid system : an assay" TRENDS IN GENETICS, vol. 10, no. 8, août 1994, pages 286-292, XP000647708 cité dans la demande voir le document en entier ---	1-10
A	LITTLE ET AL.: "Deletions within a hinge region" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE USA, vol. 82, avril 1985, WASHINGTON DC, pages 2301-2305, XP000647703 cité dans la demande voir abrégé ---	1
A	OERTEL-BUCHHEIT ET AL.: "Genetic analysis of the LexA repressor:" MOLECULAR & GENERAL GENETICS, vol. 223, 1990, BERLIN, pages 40-48, XP000647715 cité dans la demande voir abrégé -----	1

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demar internationale No
PCT/FR 97/01133

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9410307 A	11-05-94	AU 5343994 A	24-05-94
		EP 0669976 A	06-09-95
		JP 8503128 T	09-04-96
		NO 951641 A	29-06-95
		NZ 257181 A	27-07-97

US 5283173 A	01-02-94	US 5468614 A	21-11-95
		US 5667973 A	16-09-97
